NOVEL MOLECULAR MECHANISMS OF SIGNAL SENSING AND TRANSDUCTION IN BACTERIAL CHEMOTAXIS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Alise Renee Muok

May 2018

Novel molecular mechanisms of signal sensing and transduction in bacterial chemotaxis

Alise Renee Muok, Ph. D.

Cornell University 2018

Bacterial chemotaxis is a behavior bacteria exhibit to alter their motility in response to their chemical environment. The underlying sensory pathway responsible for this movement involves transmembrane receptors called methyl accepting chemotaxis proteins (MCPs) that couple with the intracellular histidine kinase CheA and adaptor protein CheW. Bacterial chemotaxis has long served as a model system for understanding the mechanisms of signal transduction. However, the exact mechanisms for the remarkable sensitivity and signal amplification of this system remains to be fully understood.

In this dissertation, I present structural and activities studies of various proteins involved in bacterial chemotaxis. In particular, I focus on elucidating the structural changes that occur in the dimeric histidine kinase CheA when it switches between the kinase-on and kinase-off activity states. In this work I have utilized a newly developed *in vitro* chemotaxis system that consists of soluble receptor mimetics that couple to CheA and CheW to form ternary complexes suitable for single-particle experiments. Electron paramagnetic resonance (EPR) and mass spectrometry experiments demonstrate that the deactivated kinase forms a more compact and rigid structure. Additionally, crystallography and cross-linking experiments demonstrate that domains in the dimer self-associate when CheA is deactivated.

In addition, I have determined the role of a previously uncharacterized protein whose gene is often found next to those for cytoplasmic receptors. We have named this new chemotaxis protein ODP1 for Oxygen Di-iron Protein and I have extensively characterized homologs of ODP1 from *Treponema denticola* and *Thermotoga maritima* using structural and biochemical methods. Analytical and spectrometric techniques demonstrate that iron and

oxygen are ligands of ODP1. Furthermore, *in vivo* assays with *T. denticola* confirm that ODP1 acts as an iron and oxygen sensor for the chemotaxis system. Crystal structures of both Td and Tm ODP1 suggests a mechanism for iron sensing.

Finally, I have developed a new method for spin-labeling proteins for EPR experiments. Specifically, I have synthesized and isolated an analog of ADP (ADP-β-S-SL) that is covalently attached to a nitroxide spin label and can hence be reconstituted into ATP-binding proteins, such as CheA. Indeed, Double Electron Electron Resonance spectroscopy experiments on the CheA kinase confirm that ADP-β-S-SL is an appropriate method for determining distance distributions from non-covalently bound spin probes.

In summary, my work has provided new insights into the domain rearrangements that occur in CheA due to receptor mediated modulation. Additionally, I've characterized a new chemotaxis sensor protein that is responsible for monitoring intracellular levels of iron and oxygen.

BIOGRAPHICAL SKETCH

Alise Muok was born on August 18th, 1988 to Lisa Guerrero and Rick Muok in Sacramento, California. She grew up in Stockton, California with her siblings, Lexi Muok, Ryan Muok and Phoenix Guerrero. Alise attended Madison Elementary School, Webster Middle School, Weber High School and eventually San Joaquin Delta College where her interest and study of biochemistry began. Prior to her studies at San Joaquin Delta College, Alise had never taken a chemistry course but thanks to Dr. Christopher Kim, she was not only able to learn chemistry, but also helped to create a lab manual and set up course equipment for the first analytical chemistry course at Delta College. After San Joaquin Delta College, Alise continued her studies in biochemistry at the University of California, Davis. At Davis, Alise worked as an undergraduate researcher in Dr. Daniel Kliebenstein's laboratory to help determine how gene combinations in *A. thaliana* affect the synthesis of secondary metabolites needed for survival. After studying at Davis for 2 years, Alise started graduate school at Cornell in 2012. In the Crane lab, Alise studies the structure and function of proteins involved in bacterial chemotaxis. Outside of lab, Alise enjoys painting, dancing and spending time outdoors. Some of her favorite outdoor activities include running, hiking and paddle boarding.

Dedicated to my brother Phoenix Guerrero and my nephew Atlas Breedlove

ACKNOWLEDGEMENTS

I would like to thank my parents and grandparents who helped support me during the many years spent pursuing my degrees. I would also like to thank my many siblings and cousins for constantly cheering me on and encouraging me to pursue my dreams. I hope that my work has not only set a good example for them but has also shown them that they can obtain whatever they want if they're not willing to give up.

I couldn't have accomplished any of my research without the help of my lab mates. Not only have all of them at some point helped me with my work, but they have also provided emotional support during difficult times. I feel very lucky to have worked with such friendly, supportive, interesting, and funny people during my graduate school experience.

I am also very thankful for the support and guidance from my advisor Dr. Brian Crane. He has always remained calm, friendly and patient with me during difficult times in my research, and has set the best example for what an academic advisor should be. Additionally, my committee members, Dr. Gerald Feigenson and Dr. Yuxin Mao, have encouraged me during my time at Cornell, which has helped me gain confidence in my abilities.

Finally, I would like to thank Cornell University and the Biochemistry, Molecular and Cellular Biology department for accepting me into the graduate school program and giving me the opportunity to conduct scientific research and learn new techniques at such a prestigious university.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter One: Introduction

1.1 The bacterial chemotaxis system

Bacterial chemotaxis is a sensory system used by most motile bacteria to control their movement in response to changes in the chemical environment. Underlying the high sensitivity and wide dynamic range of the chemotaxis system is a complex sensory apparatus. Chemotaxis has been well studied in the model organism *E. coli*, where its core proteins are anchored to the cytoplasmic membrane by methyl-accepting chemotaxis proteins (MCPs). MCPs are helical transmembrane proteins with long cytoplasmic domains that cluster together with the histidine kinase CheA, and the coupling protein CheW (Fig. 1A). MCPs assemble as trimers-of-dimers that further group into large, hexagonal arrays with CheA and CheW in a cytoplasmic plate at the intracellular receptor tips[1]. MCPs bind external ligands directly (attractants and repellents) and through periplasmic binding proteins. The receptors transduce the binding signals across the cellular membrane to regulate the autophosphorylation activity of CheA. CheA is composed of five domains (P1-P5) each with a distinct function, and once activated, the kinase initiates an intra-cellular phosphor-relay that ultimately controls flagellar rotation. For the *E. coli* system, an increase in repellent or decrease in attractant induces CheA autophosphorylation. Phosphorylated CheA interacts with its response regulator CheY and activates it by phosphoryl transfer. Phosphorylated CheY (CheY-P) binds directly to the flagellar rotor to cause clock-wise (CW) rotation of the flagella, disassembly of the flagellar bundle, and cell tumbling. When CheA is deactivated, the absence of CheY-P causes counter clock-wise (CCW) rotation of the rotor, bundling of the flagellar filaments, and smooth straight swimming. By alternating between tumbling and smooth swimming the cell can track gradient of attractants and repellents through a biased random walk (Fig. 1B).

Bacterial chemotaxis has long served as a model system for understanding the mechanisms of signal transduction. Although the chemoreceptor array consists of only three main components, complex protein arrangements and interactions create emergent properties that remain to be fully understood. Firstly, this system is exceptionally sensitive: a < 1% change in ligand concentration can be sensed by the system and amplification of the signal can result in a 50% change in cell tumbling bias and has a large dynamic range. Secondly the system has a large dynamic range: bacteria can to respond to ligand concentrations between nanomolar and millimolar concentrations[2]. Thirdly, there is a large signal gain across the membrane, with one receptor having the ability to influence > 20 kinase modules. The gain arises from highly cooperative interactions among component proteins[3]. Presently, it is known that interactions among receptors and between CheA dimers are partly responsible for these phenomena but the exact mechanism for system cooperativity remains elusive[4].

Many pathogenic bacteria are known to rely on the high sensitivity of the chemotaxis system to properly navigate host environments for survival[5]. Examples of such organisms are *Vibrio cholerae* and *Burrelia burgdorferi,* the pathogens responsible for cholerae and Lyme's disease, respectively. Because homologs of CheA are not found in humans, targeting the chemotaxis system through CheA-directed antibiotics offers a unique strategy for treating such diseases. Furthermore, understanding how receptors regulate CheA may provide the foundation to apply the chemotaxis system in biosensor technologies. For example, one could envisage a system where receptors are engineered to recognize a particular molecule of interest and transduce a binding signal through a synthetic or reconstituted membrane to increase enzymatic activity or produce a fluorescent signal on the other side. To engineer suitable gain and sensitivity into the system a detailed knowledge of how conformational events are propagated is needed.

1.2 Receptor structure, dynamics, and regulation

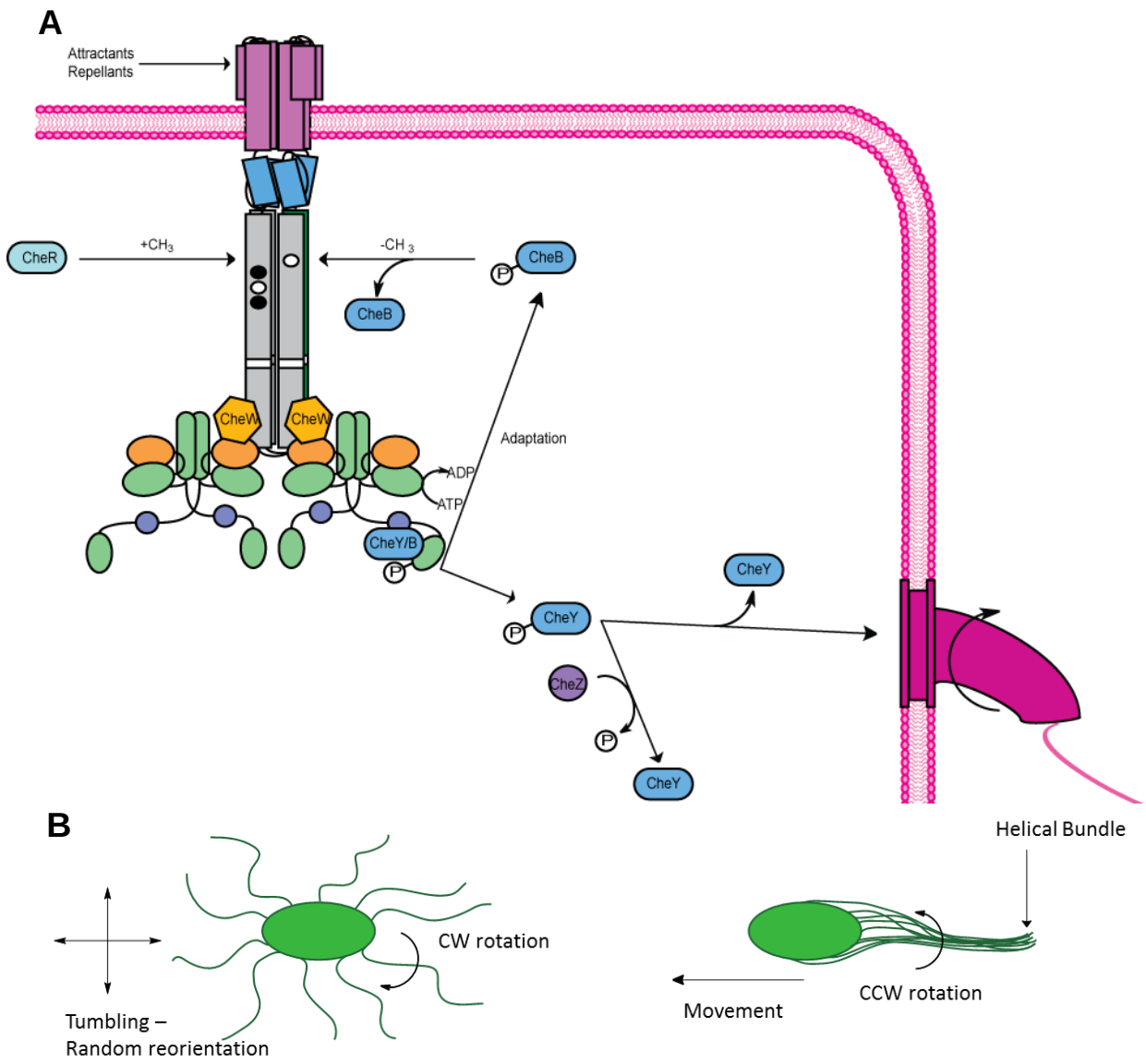


Figure 1: (A) The bacterial chemotaxis system in *Escherichia coli*. Transmembrane receptors, here depicted as a receptor dimer, bind ligands in the extracellular domain and transduce these binding events into the cell to modulate the activity of the CheA (P1-P5). CheA activates its response regulator CheY via phosphoryl transfer. Activated CheY will interact with flagellar motor to induce clock-wise (CW) rotation of the rotor. (B) *E. coli* cells switch from a state of smooth swimming to cell tumbling via disassembly of the flagellar bundle.

In general, chemoreceptors possess a canonical three-dimensional structure and domain arrangement. MCPs form dimers that are completely helical and comprise a four-helix bundle

that is ~380 Å in length. Most chemoreceptors, including the *E. coli* receptors Tar and Tsr, contain extracellular, transmembrane and intracellular domains. The extracellular domains bind attractants and repellents and undergo conformational changes upon ligand binding to transduce binding events into the cell. The intracellular domain consists of defined regions that have different roles in regulating receptor and CheA activity (Fig. 2). The intracellular region most proximal to the cell membrane is the HAMP (**H**istidine kinases, **A**denylyl cyclases, **M**ethyl-accepting chemotaxis proteins, and **P**hosphatases) domain that undergoes domain rearrangements upon ligand binding that are transmitted to the receptor tip. C-terminal to the HAMP domain are residues that undergo reversible methylation as part of an adaptation process that allows cells to keep a temporal record of ligand concentration. Glutamate residues in this region are methylated and demethylated by the proteins CheR and CheB, respectively[6]. CheR binds to a conserved sequence motif at the very C-terminal end of the major receptors, from where it can operate on neighboring molecules in addition to the one it is binding. Methylation of glutamate residues by CheR induces the receptor to an activated (kinase-on) conformation. A glycine hinge, adjacent to the methylation region, undergoes changes in local conformation to facilitate interactions among receptors that allow for cooperative modulation of CheA. The region at the very intracellular tips of receptors, called the protein interaction region (PIR), directly interacts with CheA and CheW to form the chemoreceptor ternary complex. The PIR is also where receptors interact with each other to generate the trimer-of-dimers oligomeric state.

Comprehensive mutational analyses of residues in these regions have led to detailed information about the mechanism for signal transduction. For instance, cellular studies of mutations to residues in the HAMP domain have developed models in which the HAMP helices undergo helical rotations, tilts and scissor motions to propagate signal down the length of the receptor. Furthermore, both *in vivo* and *in vitro* experiments confirm that mutating substrate

glutamate residues in the methylation region to glutamine locks them in a kinase-on state, mimicking the fully methylated receptor. These changes in methylation status correlate to local

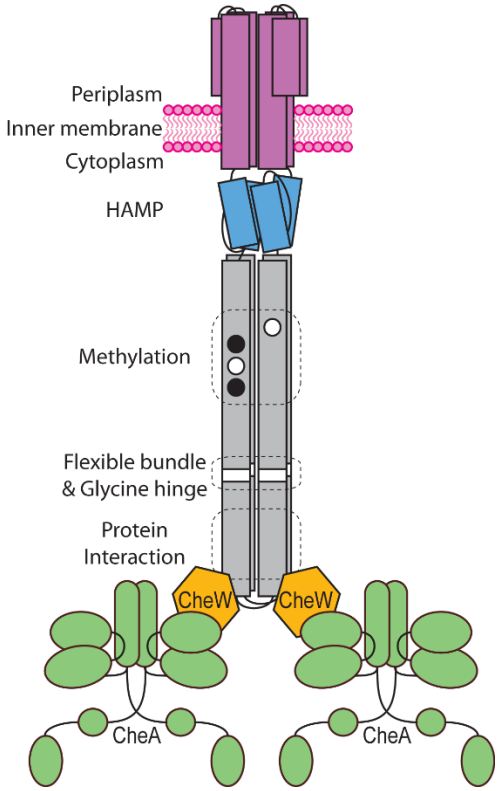


Figure 2: Specific regions of chemoreceptors carry out different roles in regulating chemotactic response. The HAMP domains transmit binding events to the receptor tips through a piston-like motion. Modification of residues in the methylation region keeps a temporal record of ligand concentration. The protein interaction region binds CheA and CheW.

dynamics near the PIR. Specifically, unmethylated receptors are ordered at the PIR (kinase-off) but this region becomes increasingly dynamic as methylation increases (kinase-on)[6]. Importantly, residue substitutions in the protein interaction region impair CheA modulation and receptor clustering.

1.3 CheA structure, dynamics, and regulation

The five domains of CheA each have a distinct role in autophosphorylation events. The P1 domain contains the substrate histidine that becomes phosphorylated, P2 interacts with the response regulator CheY, P3 is the protein dimerization domain, P4 is the Mg^{+2}-ATP binding domain, and P5 is the receptor and CheW binding domain. Autophosphorylation occurs in *trans* when the P1 domain of one subunit interacts with the P4 domain of the other subunit. The P1

and P2 domains are separated from the P3-P5 domains by long flexible linkers. The highly dynamic nature of the P1 and P2 domains makes structural characterization of full-length CheA difficult. Although high-resolution crystal structures of each domain have been determined, there is no structure of full-length CheA. Therefore, it is unclear what inter-domain interactions occur in CheA and how these interactions may differ in the kinase-on compared to the kinase-off state.

Site-directed mutagenesis of residues adjacent to the histidine substrate in P1 indicate that phosphorylation depends on a local hydrogen bonding network. In the *T. maritima* enzyme, residues Lys48, His64 and Glu67 elevates the pKa of the substrate His45 to 6.9 and stabilizes the $N^{\delta 1}$ tautomer that is necessary for CheA autophosphorylation[7]. This tautomer is unstable for solvent-exposed imidazole groups that have a pKa of 5.9. With the histidine substrate in this active tautomer, the P1 domain interacts directly with the P4 kinase domain so that the nucleophillic $N^{\varepsilon 2}$ atom can attack the γ-phosphate of ATP[7]. Although the mechanism of P1 autophosphorylation has been elucidated, the docking interaction(s) between P1 and P4 has yet to be identified and hence the CheA autophosphorylation reaction remains to be fully understood.

Several *in vivo* and *in vitro* experiments have established that there are sites on P1 that participate in direct protein-protein interactions with CheA P4, CheY and CheB[897]. Separation of the P1 domain from the kinase core (P3P4P5) demonstrates that a covalent connection to P1 is not essential for autophosphorylation[8]. However, the autophosphorylation rates for the separated domains are lower than those of the full-length kinase and a high stoichiometry is needed to attain normal chemotaxis behavior. Therefore, the long P1 linker may be important for increasing the local concentration of P1 to the kinase core or may actively steer P1:P4 interactions through long-range conformational changes. Experiments with separated P1 also

reveal that the P2 domain is not necessary for phospho transfer to the CheA response regulators CheY and CheB, but does optimize these processes for higher fidelity chemotaxis.

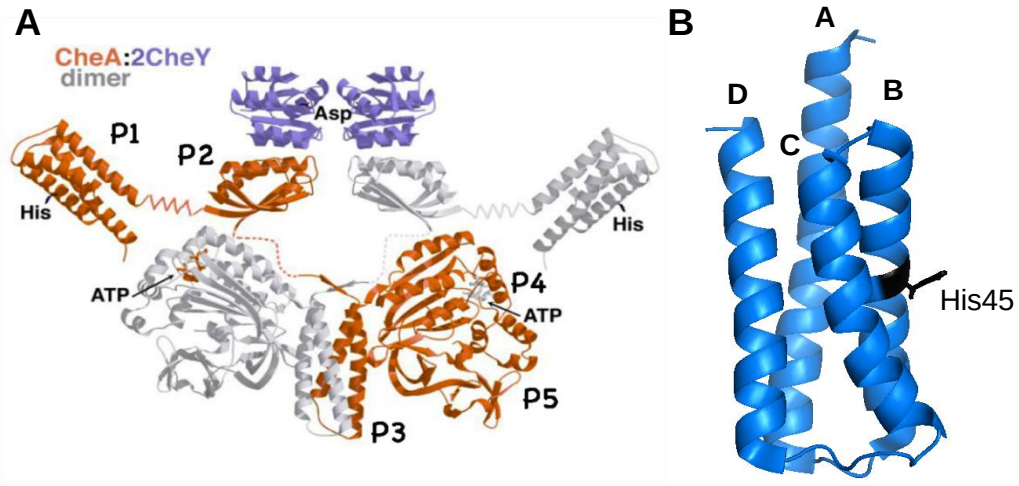


Figure 3: (A) A hypothetical full-length model of CheA using available crystal structures of truncated proteins. Autophosphorylation occurs in trans between the P1 and P4 domains. The response regulator CheY binds to the P2 domain. (B) The P1 domain of *T. maritima* CheA helices A-D. The substrate histidine His45 (black) is on helix B. Non-productive interactions with P4 localize on helix D.

The P1 domain consists of 5 helices (A-E) with A-D comprising a tightly associated four helix bundle. Helix E is located on the P1 linker and is flexible and mobile. The substrate histidine, which is residue H45 in *T. maritima* and H48 in *E. coli* CheA, is located on helix B. Mutational analyses of P1 and P4 have been conducted to identify the P1:P4 interaction site that is necessary for phosphotransfer[8]. Random mutagenesis screens of liberated P1 and P3P4P5 in *E. coli* suggest a model where residues on the helix A and B of P1 are responsible for docking interactions with the P4 domain. Indeed, this model corroborates docking simulations done with *T. maritima* protein structures and cysteine-scanning mutations that were carried out in the *Salmonella enterica* system. Intriguingly, NMR chemical shift perturbations with liberated P1 and P3P4 from *T. maritima* suggests that the strongest interaction among

these CheA fragments is at the D helix of P1, which is distal from the substrate histidine (H45), that interacts with a site on P4 that is not near the nucleotide binding region[9]. This finding suggests that there are also non-productive interactions between P1 and P4 that do not facilitate autophosphorylation. It is possible that the non-productive P1:P4 interaction is representative of a deactivated conformation of CheA that is controlled by receptors and is therefore a method for CheA modulation.

Steady-state kinetics assays that monitor CheA autophosphorylation of separated P1 from the P3P4P5 fragment of the *E. coli* kinase demonstrate that kinase activation and deactivation by nanodiscs-incorporated chemoreceptors involves changes to the catalytic rate constant (k_{cat})[10]. Furthermore, the apparent K_m for the reaction, which effectively measures the affinity of P1 to the kinase core, is not affected by receptor engagement and regulation. Therefore, these data indicates that the primary method for regulating CheA activity is the rate of phospho-transfer to P1 and not the affinity of P1 for the substrate domain P4. However, since these studies utilize liberated P1 and not the full-length kinase, additional interactions among these missing components could play a regulatory role in autophosphorylation that is not present in this system. For instance, the P1 linkers could play a role in steering P1:P4 interactions, which would effectively alter K_m values. Furthermore, as suggested by experiments discussed previously, there may be non-productive interactions among P1:P4 that don't influence the apparent K_m of the fragments but do alter productive P1:P4 binding. Importantly, the calculated K_m for the liberated P1 fragment to P3P4P5 is in the hundreds of micromolar range (100-430 μM). However, when the P1 linker is present, the effective concentration of P1 at the kinase is ~730 μM[10]. Not only do these values illustrate the importance of the P1 linker for initiating P1:P4 interactions, but demonstrate that the effective concentration is within an appropriate range of the K_m values to influence reaction rates.

Truncated variants of CheA that are missing only one of the P1 domains in the CheA dimer have elucidated additional P1 regulatory interactions. *In vitro* biochemical experiments with *T. maritima* CheA that monitor autophosphorylation demonstrate that removal of one P1 domain significantly increases CheA activity[11]. Specifically, the data suggest that the P1 domains encourage phosphatase from the P4 domains when one of them is absent. Evidence from these experiments establishes that the P1 domains possess an inhibitory function in the full-length kinase that is abrogated when one of them is removed. However, it is unclear what interactions are accountable for this inhibition. Potentially, the P1 domains are directly interacting to form a non-productive P1:P1 dimer and removal of that dimer significantly increases the activity of the full-length subunit. Alternatively, P1 could participate in non-productive interactions with other CheA domains and those contacts induce conformational changes that influence the activity of the other subunit.

Interestingly, *E. coli* possesses a naturally occurring isoform of full-length CheA ($CheA_L$), called CheA short ($CheA_S$), that does not have the first 97 n-terminal amino acids of $CheA_L$. Therefore, $CheA_S$ is missing a majority of the P1 domain including the histidine substrate (H48)[12]. *In vivo* and *in vitro* experiments have demonstrated that when $CheA_L$ and CheAs are mixed, the presence of CheAs can rescue autophosphorylation activity in defective mutants of $CheA_L$[12]. Although the mechanism for the rescue of activity is unknown, it likely involves the formation of CheA heterodimers containing only P1 domain. Hence, there are interactions between the P1 domains in the full-length CheA dimer that influence CheA activity. Notably, such interactions could influence CheA activity even when the P1 domains are free of the full-length enzyme, provided P1-P1 interactions compete with productive binding to the P4 domains; however, they would be relieved in a heterodimer, in which only P1 domain can be localized near each CheA dimer.

In addition, regulatory roles of the P4 domains have been implicated in *in vivo*, *in vitro,* and *in silico* experiments. Specifically, these studies indicate that the flexible P4 domains undergo a change in localization that may influence the accessibility of the nucleotide binding pocket to solvent or the substrate histidine on P1. Cryo-EM reconstructions of native chemoreceptor arrays composed of variant receptors that lock CheA in different degrees of activation illustrate that P4 is more localized below the P3 dimerization domain when CheA is deactivated. In support of these findings, engineered cysteine cross-linking experiments of CheA identify direct interactions between the P4 domains in the dimer[13], indicating that the mobile P4 domains can come within close proximity of each other. Furthermore, molecular dynamic simulations of CheA P3P4P5 incorporated into a chemoreceptor array reveal a dipping motion of P4 that would bring the domains within proximity of one another[14]. Although it is unclear how the motions of P4 influence CheA activity, it is apparent from the cryo-EM data that they are regulated by receptor interactions and are therefore likely a method for modulation.

Several *in vitro* experiments demonstrate that there are cooperative interactions between the subunits of a CheA dimer. *In vitro* reconstitution of a receptor:CheA:CheW complex into an artificial membrane system followed by radioisotope assays that monitor autophosphorylation in the presence of ligands demonstrates that inducing deactivation of one kinase subunit significantly reduces the activity of the other subunit[15]. Additionally, the CheA dimer exhibits negative allosteric binding of nucleotide among the P4 domains. In *T. maritima* CheA, the Kd for ATP binding in the first nucleotide pocket is ~6 μM and the Kd for the second site is ~5 mM[16]. The mechanism for communication between the P4 domains remains unclear. It's possible that direct contacts between the P4 domains facilitate these allosteric interactions. Alternatively, signals may be transmitted through the P3 dimerization domains via conformational changes at the P3-P4 linker.

Mutagenesis experiments on the linkers that connect the P4-P5 and P3-P4 domains demonstrate that specific residues in these linkers are critical for signal propagation through the kinase. Specific mutations in the P4-P5 linker can still retain WT *in vitro* kinase activity but fail to be regulated by receptors both *in vitro* and *in vivo*[17]. Therefore, these perturbations prevent transduction of P5:receptor interactions to the rest of the kinase. Interestingly, other mutations in the P4-P5 linker increase or decrease WT CheA activity even in the absence of the P5 domain[18]. These mutants are also defective in receptor-mediated modulation of activity. The fact that these linker perturbations can alter CheA activity without P5 suggests that receptors modulate kinase activity by controlling the conformation of the linker. Furthermore, mutations in the P3-P4 linker not only inhibit receptor-mediated activation but also decrease the basal autophosphorylation level of the kinase[17]. These data suggest that the distortions in the P3-P4 linker put the kinase in an unfavorable orientation that inhibits autophosphorylation. Collectively, it appears that both the P4-P5 and P3-P4 linkers are involved transmitting and maintaining signals from receptors to the kinase domain.

1.4 The adaptor protein CheW

The last integral component of the chemoreceptor array is the adaptor protein CheW that interacts directly with receptors and CheA. Although CheW itself does not possess any enzymatic activity, it is needed for proper modulation of CheA activity by receptors. Crystal structures of CheW and CheA P5 reveal that they are structural paralogs—both consist of two perpendicular β-barrels with a hydrophobic core in between them. A crystal structure consisting of a receptor cytoplasmic fragment in complex with the CheA P5 domain and CheW (3UR1) reveals the interactions of CheW with its binding partners for maintaining array formation and propagating signals across the array[19]. Interestingly, the P5 domain and CheW conserve hydrophobic surfaces at each end of their intertwined β-barrels and it is these interfaces that mediate contacts to form hexagonal rings composed of alternating P5/CheW subunits[19,20]. The

receptor binds P5 and CheW on the outside of these rings at a groove between the β-barrels of each protein. The presence of P5/CheW rings that bridge receptors suggests that CheW may play a mechanistic role in transmitting cooperativity among receptors. Indeed, *in vivo* mutational experiments confirm that disruptions in the CheW protein:protein interaction regions produce arrays with aberrant cooperative behavior[21]. Therefore, the role of CheW in chemotaxis is not strictly architectural.

1.5 In vivo arrangement of chemoreceptor arrays

Although chemoreceptors are often depicted as functioning as a receptor dimer, they function *in vivo* as a trimer-of-receptor dimers that further assemble into a large hexagonal supramolecular array[1]. The hexagonal arrangement of the receptor trimers-of-dimers is upheld by interactions among CheA P5 and CheW at the receptor tips, and among the tips themselves. Interestingly, *in vivo* cryo-electron tomography maps generated from whole bacterial cells demonstrate that this receptor arrangement is conserved in all classes of chemotactic bacteria studied[1]. Protein crystal structures generated using fragments of receptors, two domains of CheA (P4P5), and CheW allowed for interpretation of *in vivo* cryo-EM maps at the atomic level (3UR1)[1,19]. However, several of the domains of CheA (P1,P2,P4) and portions of the chemoreceptors do not have clearly discernible density in these maps due to their dynamic nature. Therefore, it is unclear what conformational changes occur in the array to initiate CheA on-off switching.

In order to determine the placement of the mobile CheA domains in the *in vivo* arrays, cells devoid of an adaptation system (-CheR/-CheB) were used in conjunction with chemoreceptors that are either unmethylated (4E, CheA-off) or mutated methylation mimics (4Q, CheA-on)[22]. Although the P1, P2 and P4 domains could not be accurately placed in the resulting cryo-electron tomography (cryo-ET) density maps, there is a significant increase in electron density below the P5/CheW rings in the 4E, CheA-off tomograms. This data supports a model where

the P1, P2 and P4 domains become less mobile and localize near the P4 domain in the deactivated state.

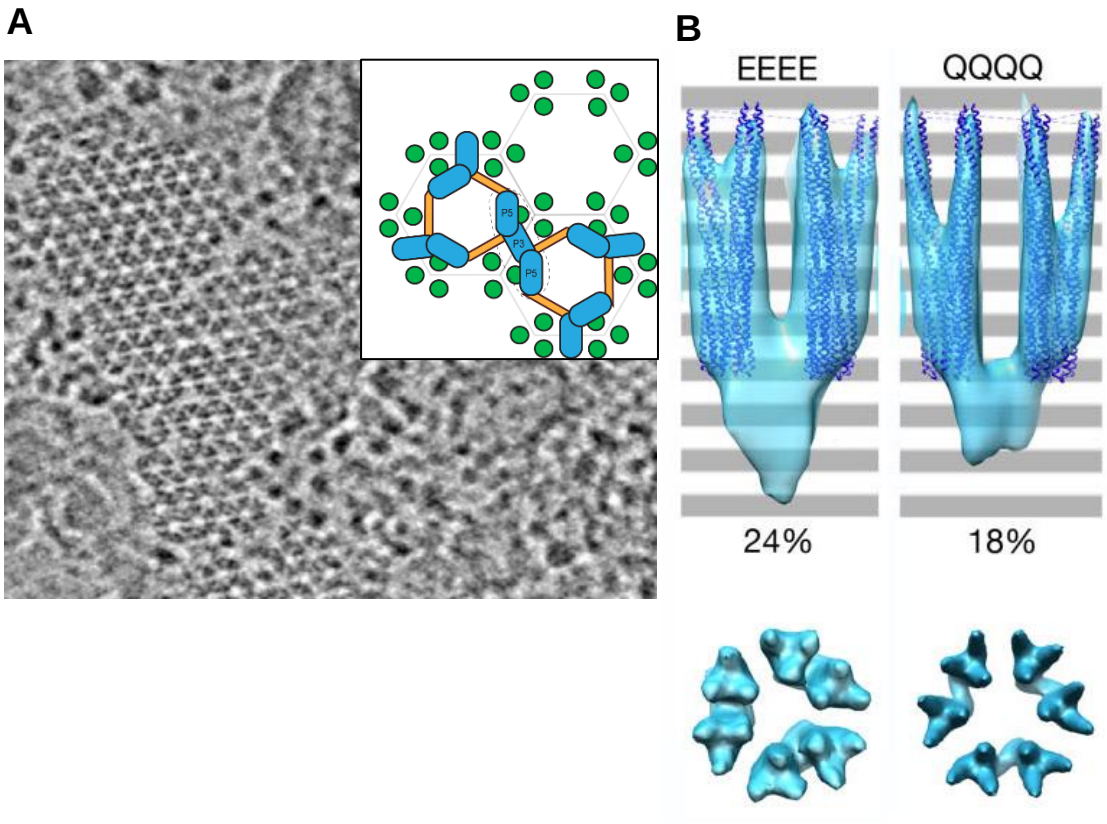


Figure 4: (A) Whole cell micrographs reveal the hexagonal pattern of the chemoreceptor array. Inset shows a cartoon representation of protein arrangements in the array. Image from Briegel et al. 2012 (B) Averaged tomographs of *in vivo* arrays with receptors that are either completely unmethylated (EEEE, kinase-off) or are mutated to mimic fully methylated receptors (QQQQ, kinase-on) demonstrate that electron density below the receptors increases in the kinase-off state. Image from Briegel et al. 2010.

Presently, cryo-ET experiments have generated density maps that are 11.3 Å resolution[14]. To determine the conformational changes that occur in CheA at a higher resolution, attempts at producing three-dimensional single-particle reconstructions (SPA) of the ternary complexes via cryo- electron microscopy (cryo-EM) are currently being pursued by several research groups. *In*

vitro experiments using nano-disc incorporated receptors have demonstrated that the smallest functional unit of the chemoreceptor array consists of two trimer-of-dimers with one CheA dimer and two CheW subunits. In this unit, referred to as the chemotaxis 'core' unit, each receptor trimer-of-dimer directly interacts with one CheA P5 domain and one CheW[23]. When this core unit is reconstituted *in vitro*, full modulation of CheA activity is exhibited. Therefore, to accurately determine changes that occur in the complex between the kinase-off and kinase-on states using SPA, an *in vitro* system that consists of the chemotaxis core is ideal.

1.6 Generation of soluble receptor mimetics

As seen in *in vivo* maps as well as *in vitro* experiments, receptor dimer trimerization is critical for proper CheA modulation. To adequately conduct *in vitro* structural experiments with receptor:CheA:CheW complexes, a trimers-of-dimers receptor state is required. However, establishing such a system is taxing due to the insoluble nature of transmembrane receptors. There are methods to reconstitute receptors into artificial membranes but these methods do not provide directional selectivity for receptor insertion; reconstitution can produce antiparallel receptors. However, when only the soluble kinase control module (KCM) portions of receptors are used in *in vitro* experiments, they can bind to and regulate CheA activity if they are prompted to form the trimer-of-dimer state by lipid templating or the presence of a stabilizing agent.

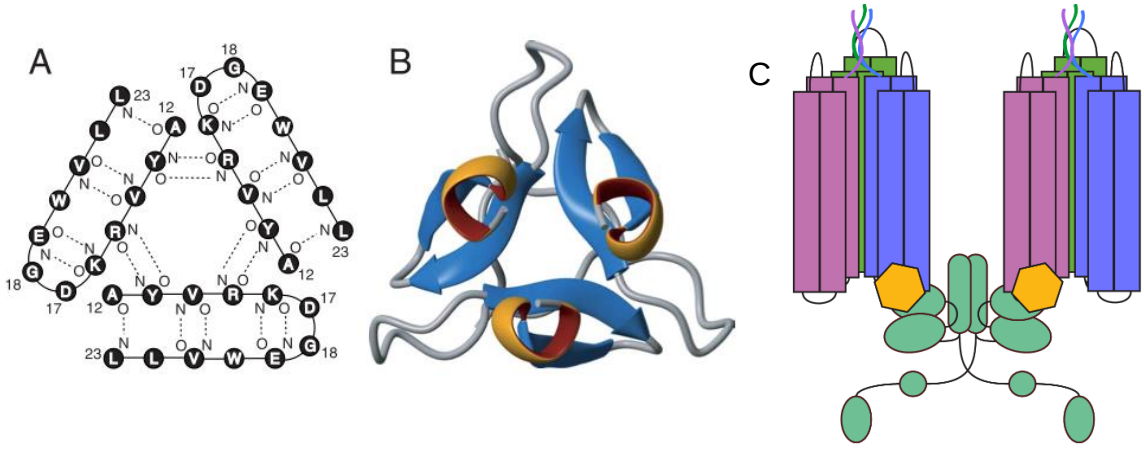


Figure 5: (A) The residues of the trimerization motif of the T4 phage protein fibritin. (B) Each subunit in the trimerization motif forms a β-hairpin that interacts with two other subunits to form a trimeric β-propeller. (C) Single-chain receptor dimers, where two subunits were joined by a protein linker, were covalently attached to the trimerization motif to pre-form a trimer-of-dimer arrangement.

To generate a suitable soluble system that can produce receptor:CheA:CheW complexes with the receptors in the trimerized state, Greenswag et al. engineered receptor mimetics using cytoplasmic fragments of receptors that pre-form a trimer-of-dimer state[24]. Specifically, various lengths of KCMs from the *E. coli* Tar receptor and *T. maritima* Tm14 receptor were first fused together to form a connected receptor dimer and then capped with a trimerization motif that is natively found in the T4 phage protein fibritin (Fig 5 A,B). This trimerization motif consists of 12 residues in a β-hairpin that form a β-propeller trimer upon expression. Likewise, the engineered receptor chimeras form a trimer after purification. Further characterization of the variant receptors demonstrate that they strongly interact with CheA and CheW, and modulate CheA activity to a greater extent than receptor dimers that do not possess the trimerization motif.

1.7 Soluble receptors without transmembrane domains

Some species of bacteria, such as *Thermotoga maritima* and *Bacillus subtilis*, are known to possess cytoplasmic receptors that do not contain a transmembrane region or an obvious sensor

domain. Although cytoplasmic receptors do not have transmembrane domains, they assemble into arrays similar to transmembrane receptors[19]. Like membrane-bound receptors, cytoplasmic receptors function as a trimer-of-dimers and further organize into a hexagonal lattice bound to a baseplate of CheA and CheW at their signaling tips. While membrane arrays are stabilized by interactions with the cellular membrane, the cytoplasmic lattice is stabilized by molecular crowding and by interactions among receptors at their terminal domains—thus creating a 'sandwich' of receptors flanked by two CheA/CheW baseplates (Fig. 6). The middle of the sandwich, which contains the putative signaling domains of the receptors, is porous enough to allow small molecules to freely diffuse in the array. Due to similarities in organization between cytoplasmic and membrane-bound receptors, an *in vitro* system consisting of cytoplasmic receptors will likely be useful for understanding the interactions present in transmembrane systems.

One identified cytoplasmic receptor is TM14 from *T. maritima*. The structure of TM14 reveals the classic coil-coil dimeric structure of MCPs[25]. Like transmembrane MCPs, TM14 has an adaptation region that contains sites for modification, and a signaling tip that interacts with CheA and CheW. Radioisotope assays monitoring CheA activity demonstrate that unmodified TM14 deactivates CheA *in vitro*. Although TM14 is known to be a receptor that influences CheA activity, its role in chemotaxis remains to be understood. There is evidence that cytoplasmic receptors respond to the metabolic state of the cell, but it is unknown what signals or ligands the

receptors recognize. Furthermore, cytoplasmic receptors have no obvious sensor domain, so their sensory mechanisms are largely a mystery.

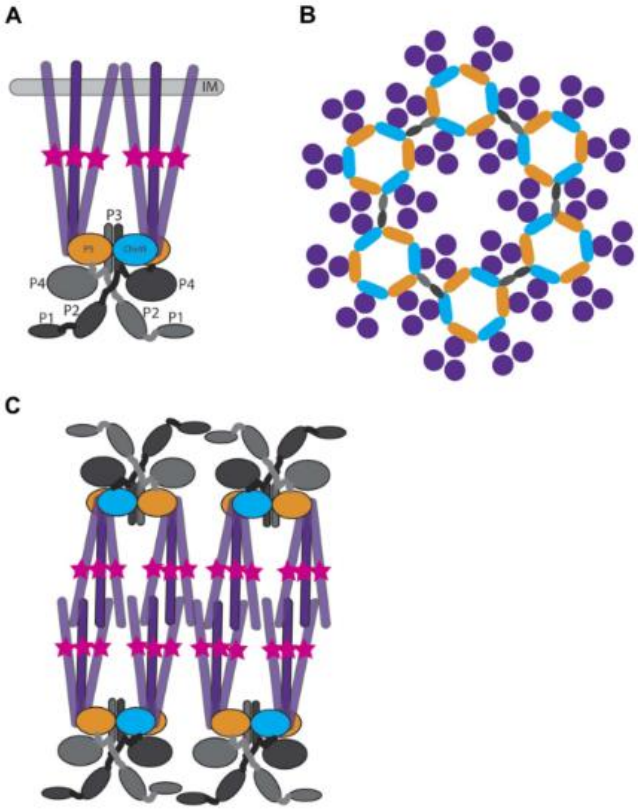


Figure 6: (A) Transmembrane receptors form a trimer-of-dimers that interact with CheA (orange and grey) and CheW (blue) on their cytoplasmic tips, and (B) further organize into hexagonal arrays. (C) Cytoplasmic receptors form a sandwich of receptor trimer-of-dimers, each side of the sandwich forming identical hexagonal arrays of transmembrane receptors. Image from Briegel et al. 2014.

1.8 References

1. Briegel, a. *et al.* Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. *Proc. Natl. Acad. Sci.* **109,** 3766–3771 (2012).

2. Sourjik, V. & Wingreen, N. Responding to Chemical Gradients : Bacterial Chemotaxis. **24,** 262–268 (2013).

3. Falke, J. J. Cooperativity between bacterial chemotaxis receptors. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 6530–2 (2002).

4. Asinas, A. E. & Weis, R. M. Competitive and cooperative interactions in receptor signaling complexes. *J. Biol. Chem.* **281,** 30512–23 (2006).

5. Miguel A Matilla, T. K. The effect of bacterial chemotaxis on host infection and pathogenicity. *FEMS Microbiol. Rev.* **Volume 42,**

6. Mello, B. a, Shaw, L. & Tu, Y. Effects of receptor interaction in bacterial chemotaxis. *Biophys. J.* **87,** 1578–95 (2004).

7. Quezada, C. M. *et al.* Structural and chemical requirements for histidine phosphorylation by the chemotaxis kinase CheA. *J. Biol. Chem.* **280,** 30581–5 (2005).

8. Nishiyama, S. I., Garzón, A. & Parkinson, J. S. Mutational analysis of the P1 phosphorylation domain in escherichia coli CheA, the signaling kinase for chemotaxis. *J. Bacteriol.* **196,** 257–264 (2014).

9. Hamel, D. J., Zhou, H., Starich, M. R., Byrd, R. A. & Dahlquist, F. W. Chemical-shift-perturbation mapping of the phosphotransfer and catalytic domain interaction in the histidine autokinase CheA from Thermotoga maritima. *Biochemistry* **45,** 9509–9517 (2006).

10. Pan, W., Dahlquist, F. W. & Hazelbauer, G. L. Signaling complexes control the chemotaxis kinase by altering its apparent rate constant of autophosphorylation. *Protein Sci.* **26,** 1535–1546 (2017).

11. Greenswag, A., Muok, A., Li, X. & Crane, B. HHS Public Access. **427,** 87–92 (2016).

12. Wolfe, a J. & Stewart, R. C. The short form of the CheA protein restores kinase activity and chemotactic ability to kinase-deficient mutants. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 1518–22 (1993).

13. Careaga, C. L. & Falke, J. J. Structure and dynamics of Escherichia coli chemosensory receptors. Engineered sulfhydryl studies. *Biophys. J.* **62,** 209-16; discussion 217–9 (1992).

14. Cassidy, C. K. *et al.* CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. *Elife* **4,** 1–20 (2015).

15. Li, M. & Hazelbauer, G. L. Selective allosteric coupling in core chemotaxis signaling complexes. *Proc. Natl. Acad. Sci.* **111,** 15940–15945 (2014).

16. Eaton, A. K. & Stewart, R. C. NIH Public Access. **48,** 6412–6422 (2010).

17. Wang, X., Wu, C., Vu, A., Shea, J. E. & Dahlquist, F. W. Computational and experimental analyses reveal the essential roles of interdomain linkers in the biological function of chemotaxis histidine kinase CheA. *J. Am. Chem. Soc.* **134,** 16107–16110 (2012).

18. Ding, X., He, Q., Shen, F., Dahlquist, F. W. & Wang, X. The Regulatory Role of an Interdomain Linker in the Bacterial Chemotaxis Histidine Kinase CheA. *J. Bacteriol.* JB.00052-18 (2018). doi:10.1128/JB.00052-18

19. Briegel, A. *et al.* Structure of bacterial cytoplasmic chemoreceptor arrays and implications for chemotactic signaling. 1–16 (2014). doi:10.7554/eLife.02151

20. Bhatnagar, J. *et al.* Structure of the ternary complex formed by a chemotaxis receptor signaling domain, the CheA histidine kinase, and the coupling protein CheW as determined by pulsed dipolar ESR spectroscopy. *Biochemistry* **49,** 3824–41 (2010).

21. Pedetta, A. *et al.* HHS Public Access. **93,** 1144–1155 (2015).

22. Briegel, A. *et al.* NIH Public Access. **14,** 384–399 (2010).

23. Boldog, T., Grimme, S., Li, M., Sligar, S. G. & Hazelbauer, G. L. Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11509–11514 (2006).

24. Greenswag, A. R. *et al.* Preformed Soluble Chemoreceptor Trimers That Mimic Cellular Assembly States and Activate CheA Autophosphorylation. *Biochemistry* **54,** 3454–68 (2015).

25. Pollard, A., Bilwes, A. & Crane, B. NIH Public Access. **40,** 1301–1315 (2012).

818792414161517191231074620

Chapter Two

The mechanism of CheA kinase activation in chemoreceptor arrays

2.1 Introduction

2.1.1 Receptor clustering and array formation

To sense and respond to changes in the chemical environment, many bacteria possess a sensory apparatus in their cytoplasmic membranes composed of methyl-accepting chemotaxis proteins (MCPs), the histidine kinase CheA and the coupling protein CheW. MCPs are transmembrane receptors that form trimers-of-dimers and cluster into large, hexagonal arrays (Fig. 1)[1]. These receptor clusters have remarkable properties in terms of signal gain, sensitivity and dynamic range. The transmembrane domains of MCPs anchor the array to the membrane while CheA and CheW form a cytoplasmic plate at the intracellular receptor tips. MCPs bind external ligands and transduce signals across the cellular membrane to regulate the autophosphorylation activity of CheA. The smallest unit of the receptor array that maintains full modulation of CheA consists of two receptor trimer-of-dimers, one CheA dimer and two CheWs[2]. Although it is known that MCPs modulate CheA activity, the exact mechanism for this regulation remains elusive due to the fact that the protein contains long unstructured linkers and undergoes large-scale conformational changes between activity states. Chemotaxis has long served as a key model system for understanding general aspects of transmembrane signaling. Unraveling the molecular mechanism of how the receptors engage and regulate CheA would greatly advance our current understanding of this remarkable system.

2.1.2 Soluble receptor mimetics pre-form trimer-of-dimer formation

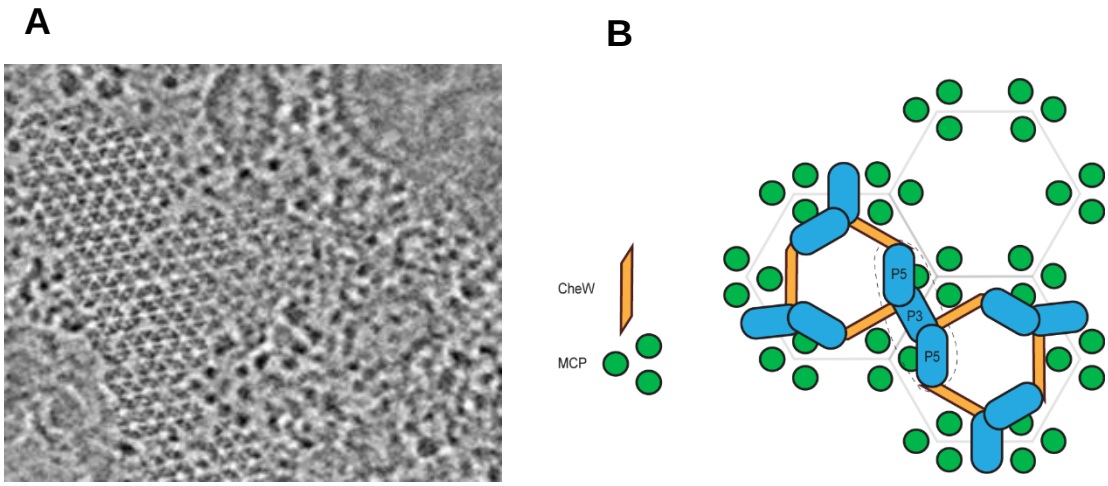


Figure 1: (A) An *in vivo* transmembrane chemoreceptor array from whole *E. coli* cells reveals the hexagonal arrangement of receptors with CheA and CheW. Image from Briegel et al. 2012. (B) A cartoon representation of the protein arrangement demonstrates that the receptors form a trimer-of-dimers linked by rings consisting of the CheA P5 domain and CheW.

Thus far, we have generated chimeric receptor variants that are soluble, pre-form a trimer-of-dimer arrangement, and associate with CheA and CheW[3]. To circumvent the complications of reconstituting full-length transmembrane receptors, the variants consist of fused single-chain receptor cytoplasmic domains linked to a “fold-on” trimerization motif from the T4 phage protein fibritin (Fig. 2A). Thus, these receptor variants form the oligomeric state necessary for CheA modulation without requiring membrane incorporation. Variants were generated based on cytoplasmic domains from the *T. maritima* MCP Tm14, and the *E. coli* MCP Tar. Hereafter the receptor variants will be referred to as the Tm14 foldon and Tar foldon. Small-angle x-ray scattering, multi-angle light scattering (MALS) and pulsed-ESR experiments confirm that the foldons trimerize *in vitro*[3]. Furthermore, pull-down assays, MALS experiments, and radioisotope assays indicate that they also retain the ability to complex with CheA and CheW, and modulate CheA activity (Fig 2B). Importantly, autophosphorylation assays confirm that the foldons lock

CheA into specific activity states, where the kinase is either active or inhibited. Structural analysis of these complexes may reveal key conformational differences in CheA, as well as the nature of interactions among components that are responsible for changes in CheA activity.

Herein, I present experiments conducted with receptor foldon:CheA:CheW ternary complexes to attempt to characterize conformational changes that occur in CheA between activity states (kinase on vs. kinase off). Our first efforts at characterizing the complexes were accomplished by mixing the composite proteins, isolating complexes via size-exclusion chromatography (SEC), and using these samples for structural experiments. To optimize the homogeneity and stability of these complexes, we then employed the chemical crosslinker DSSO in the purification procedure before SEC. Indeed, these cross-linked complexes are homogenous, stable and have provided a ternary chemotaxis complex ideal for single-particle reconstructions. Electron paramagnetic resonance (EPR) double electron-electron resonance (DEER) spectroscopy and mass spectrometry (MS) were carried out with the complexes and free CheA to differentiate CheA conformational changes.

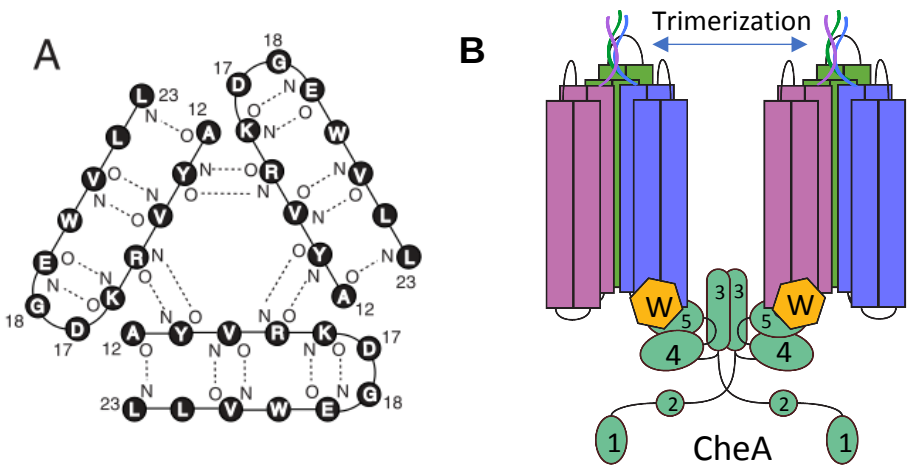


Figure 2: (A) The trimerization motif of the T4-phage protein fibritin consists of 12 residues that form a tightly associated trimer. This trimerization sequence is referred to as a foldon motif. (B) Cytoplasmic fragments of receptors were linked into dimers via short linkers and then capped with a trimerization motif to induce a pre-formed trimer-of-dimer arrangement. These receptor mimetics interact with CheA and CheW to form ternary complexes *in vitro* without a membrane component.

2.2 Results and Discussion

2.2.1 Characterization of foldons

To ensure that the Tar and Tm14 receptor foldons trimerize in solution to form the globular trimer-of-dimer formation as expected, multi-angle light scattering (MALS) and small-angle x-ray scattering (SAXS) experiments were conducted with the purified proteins[3]. As expected, the MALS data demonstrates that the foldons form the expected ~60 kDa trimer in solution (Fig 3A). Interestingly, when a sample of the Tar foldon is run on an SDS-PAGE gel, this trimer is also present and will only disassemble into ~20 kDa subunits if the sample is boiled before loading. Furthermore, SAXS experiments published in Greenswag et al. with the purified foldons demonstrate that they form the expected globular shape in solution[3].

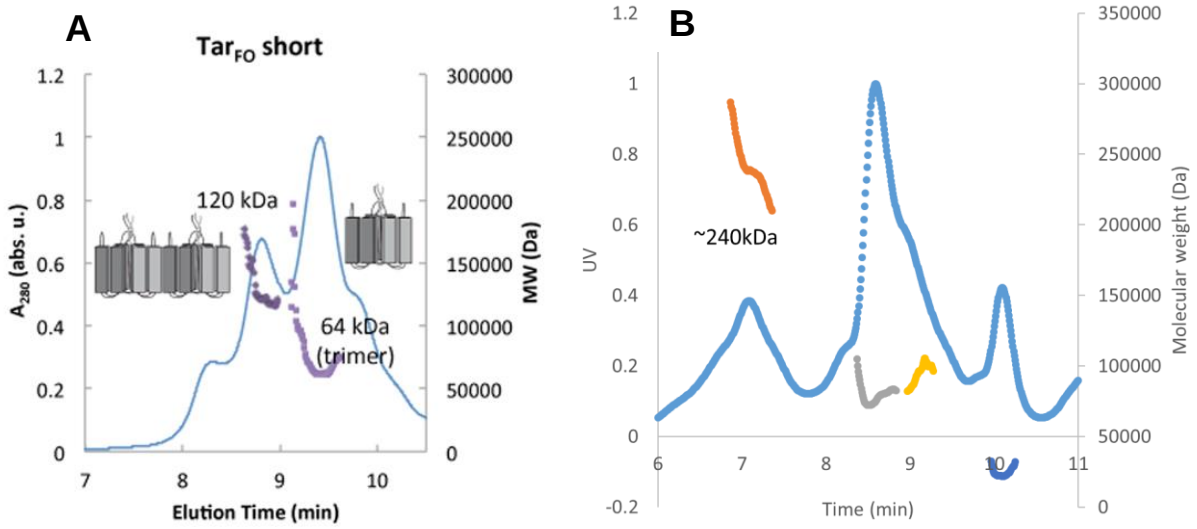
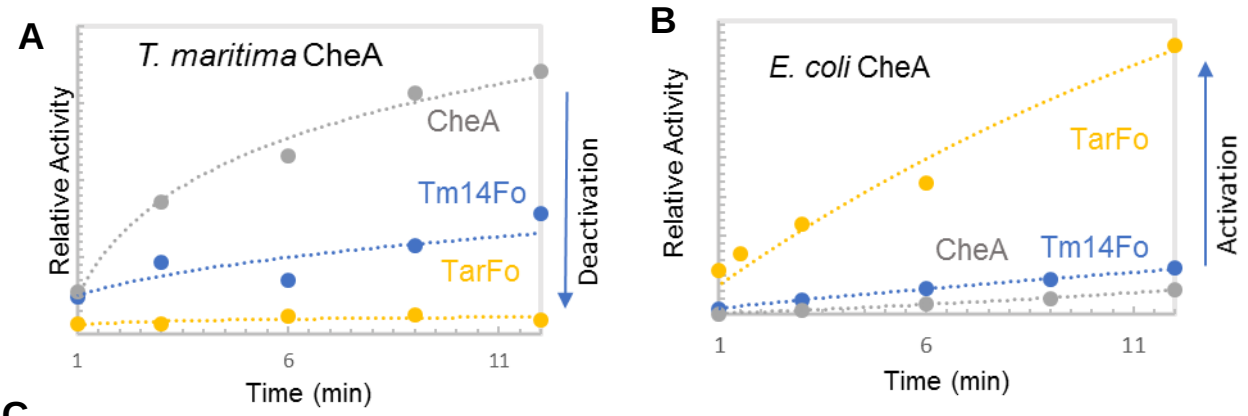


Figure 3: Multi-angle light scattering experiments demonstrate that (A) receptor foldons trimerize *in vitro* and (B) foldons incubated with CheA and CheW in a 1:1:1 ratio form ternary complexes *in vitro*. Specifically the predominant species consists of one CheA dimer, 2 CheW subunits and one receptor foldon.

MALS experiments were conducted to determine if the foldons interact with CheA and CheW in solution to form ternary chemotaxis complexes. When the foldons, CheA and CheW are mixed in a 1:1:1 ratio, a 240 kDa complex is most abundant (Fig. 3B). Based on the molecular weight of the three composite proteins, this complex consists of 1 CheA dimer, 2 CheW proteins, and a trimerized foldon. Interestingly, Greenswag et al. has shown that when purified receptor dimers (no trimerization motif) are used in identical MALS experiments, there is not a peak corresponding to a ternary complex[3]. This indicates that using the trimerization motif to pre-form a trimer-of-dimer arrangement induces the receptor affinity to CheA and CheW.

2.2.2 Receptor foldons modulate CheA autophosphorylation



C

Tm CheA Samples	Relative [CheA-P]	[CheA-P] 95% confidence bounds	Rate constant (k min^{-1})	k min^{-1} 95% confidence bounds	[CheA-P]*k
CheA	16.82	+/- 11.5	0.165	+/- 0.25	**2.71**
Tm14Fo+CheA	272.6	+/- 163	0.0016	+/- 0.94	**0.43**
TarFo+ CheA	1.86	+/- 3.5	0.214	+/- 1.14	**0.39**

D

Ec CheA Samples	Relative [CheA-P]	[CheA-P] 95% confidence bounds	Rate constant (k min^{-1})	k min^{-1} 95% confidence bounds	[CheA-P]*k
CheA	472.4	+/- 3000	0.00234	+/- 0.19	**1.1**
Tm14Fo+CheA	19.52	+/- 16.24	0.159	+/- 0.29	**3.1**
TarFo+ CheA	353.8	+/- 3000	0.0229	+/- 0.21	**8.1**

Figure 4: Radioisotope assays that monitor CheA autophosphorylation indicate that the receptor foldons (A) deactivate *T. maritima* CheA and (B) activate *E. coli* CheA. Furthermore, the fold decrease in activity for Tm CheA and fold increase for Ec CheA is identical for each receptor mimetic. (C) First order rate constants calculated from radioisotope experiments with *T. maritima* CheA and (D) *E. coli* CheA.

The ability of the receptor foldons to modulate CheA activity was determined by conducting radioisotope labeling assays with radio-labeled ATP-γ-P^{32} (Fig. 4). Interestingly, both the Tar and Tm14 receptor foldons deactivate *T. maritima* CheA and activate *E. coli* CheA. Furthermore, the foldons modulate Tm CheA and Ec CheA to the same fold decrease and increase, respectively. Specifically, the Tar foldon deactivates Tm CheA twenty-fold and

increases Ec CheA activity twenty-fold (Fig. 4A, 4C). Likewise, the Tm14 foldon deactivates TmA two-fold and activates EcA two-fold (Fig. 4B, 4D).

2.2.3 Generation of receptor:CheA:CheW complexes

In order to conduct structural experiments with the ternary complexes, they first needed to be isolated in large-scale quantities. To determine the best conditions for optimal ternary complex formation, equimolar mixtures of the three proteins were incubated in candidate buffer conditions and then examined via multi-angle light scattering to determine the molecular weight and relative abundance of ternary complexes formed (Fig. 5). Published data indicates that receptor:CheA:CheW binding increases with high salt and low pH due to neutralization of negatively charged residues that surround the protein:protein interaction regions[4]. Therefore, conditions with higher than physiological salt and lower than physiological pH were tested. Collectively, the data demonstrates that the ‘core’ complex is formed in higher abundance when

the salt concentration is increased from 150 mM to 250 mM, the pH is lowered, or if the salt is changed from NaCl to KCl.

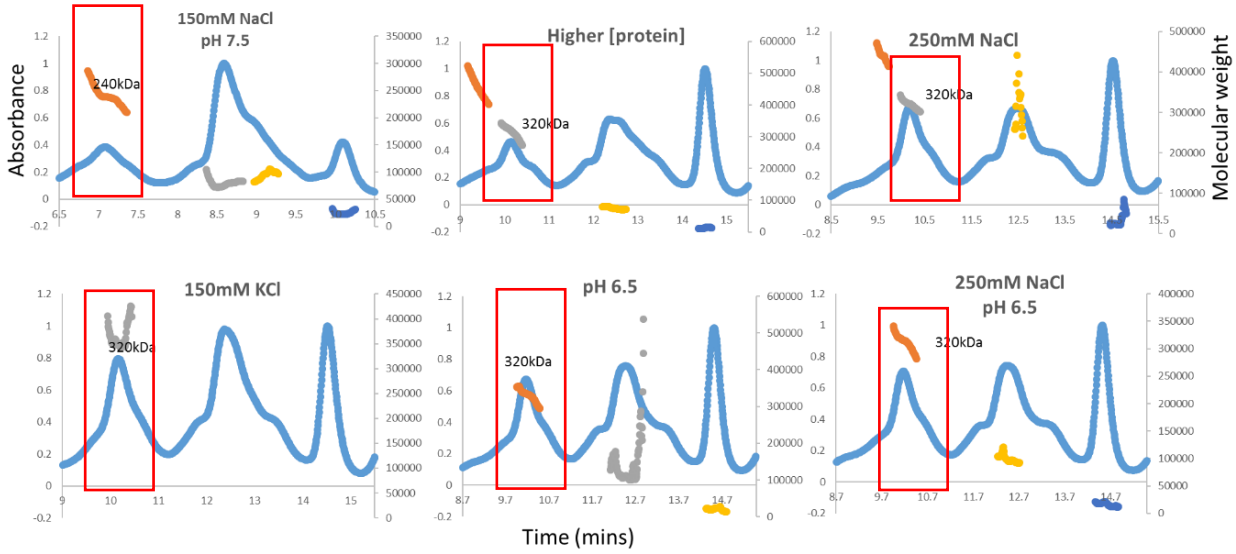


Figure 5: MALS experiments with samples of foldon, CheA and CheW mixed in a 1:1:1 ratio in different conditions demonstrates that changing buffer conditions and protein concentrations increases the average molecular weight and relative abundance of ternary complexes. Specifically, increasing protein concentration, increasing sat concentration, the addition of KCl, or lowering the buffer pH increases the molecular weight from 240 kDa to 320 kDa. Furthermore, the total UV absorbance suggests that this 320 kDa complex is more abundant under these conditions.

Based upon the data collected from MALS, large scale preparations of the complexes were prepared to conduct radioisotope, SEC-SAXS and electron-microscopy experiments. The three proteins were incubated in a high salt, low pH buffer and then run on SEC. Two apparent SEC peaks were concentrated and then ran on MALS to reveal the presence of two high molecular weight ternary species (Fig. 6). The MALS data suggests that although there is a dominant species in the two samples, they are not homogenous as the molecular weight of the

complexes decreases over the elution time. This heterogeneity is likely due to dilution effects and subunit exchange among the complexes.

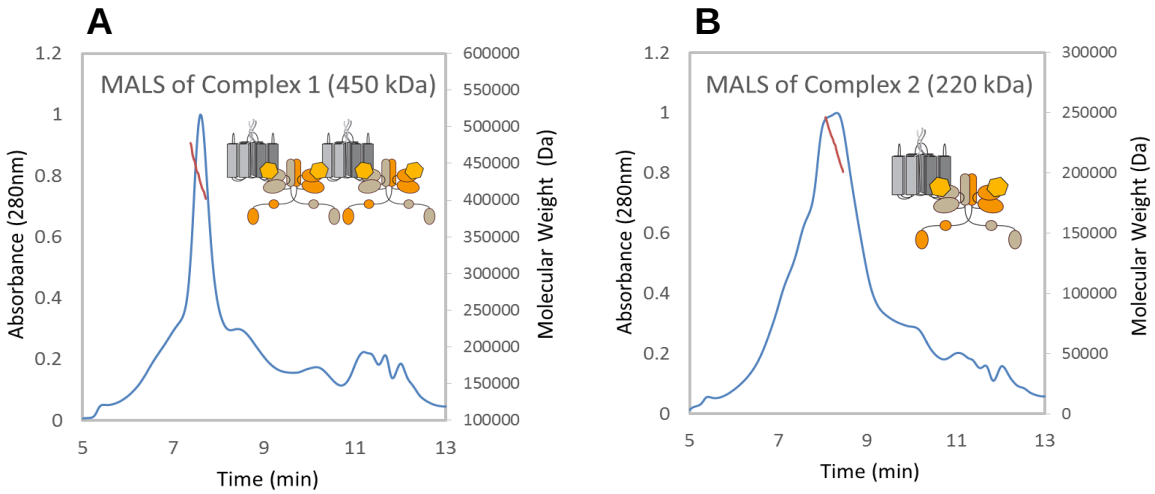


Figure 6: SEC-MALS experiments of isolated ternary complexes demonstrates that the particles are relatively stable and homogenous. The SEC trace suggests that these ternary complexes do not fall apart into composite proteins after a freeze/thaw process and dilution on the SEC. Furthermore, the molecular weight estimates suggest a dominant species present in each sample.

2.2.4 Mutation of critical residues in the receptor foldons

To ensure that native receptor-trimer contacts have been preserved in the engineered foldons, residues in the Tar foldon that have been previously identified as crucial for receptor trimerization and ternary complex formation *in vivo* were mutated. As predicted, the integrity of the foldon is compromised and can no longer trimerize or interact with CheA:CheW. Specifically, in the Tar foldon residues E391 and R388 are located at the CheA/CheW binding site and mutating these residues has been found to prevent complex formation and induce cell tumbling bias[5]. In the Tm14 foldon, residue F396 has been found to participate in pi-stacking interactions within the receptor dimer and these contacts may induce on-off switching by perturbing receptor:receptor interactions[6].

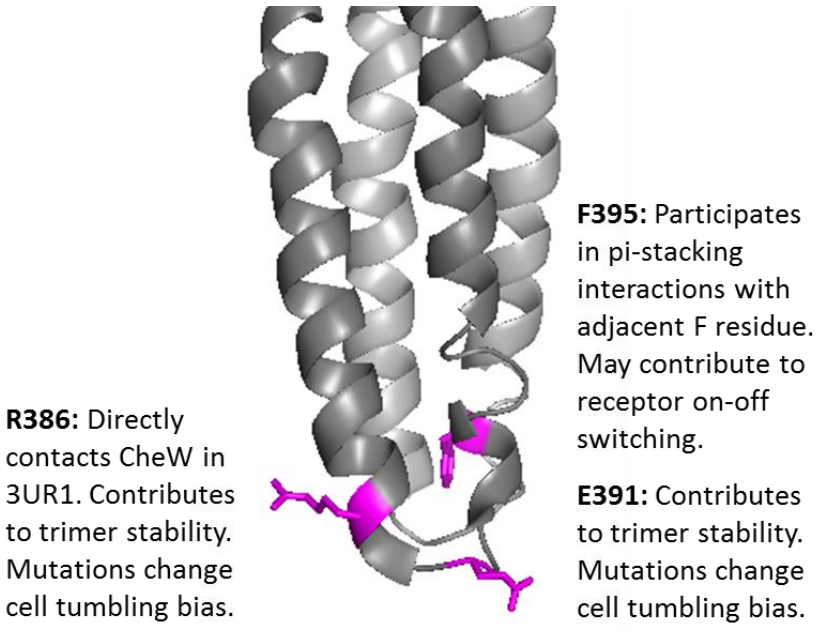


Fig 7: Residues in the receptor foldons were mutated based on previous *in vivo* mutational experiments with full-length receptors. Specifically, residues R386 and E391 are found to be imperative for proper CheA modulation and receptor clustering. Additionally, F395 participates in pi-stacking interactions that may modulate receptor conformation.

Based on published experiments, the following mutations were made in the Tar foldon: R386F, R386W, E391A (Fig. 7). Interestingly, the R388F mutant was found to be significantly more stable than the unmutated foldon. Not only was the protein purified in higher abundance, but it is stable at much higher concentrations than the unmutated foldon. However, DSSO crosslinking revealed that this mutant is mostly present as a dimer in solution and does not form a stable ternary complex with CheA and CheW. Furthermore, radioisotope assays demonstrate that the R386F mutant only exhibits around 1.5-fold deactivation of CheA; the unmutated Tar foldon induces 20-fold deactivation (Fig. 8).

For the Tar R388W mutant, MALS experiments show that it's only present as a monomer in solution. Furthermore, DSSO experiments show that it does not form a ternary complex with CheA and CheW, and radioisotope assays show that it does not modulate CheA activity. The

Tar E391A mutant was extremely unstable and is present only as high molecular-weight aggregates after purification on size-exclusion chromatography (SEC). Therefore, this mutant could not be used for *in vitro* experiments. Lastly, the Tm14 foldon was mutated to F396W to see if this mutation influences CheA on-off switching *in vitro*. However, radioassays demonstrate that the mutation does not change the ability of the foldon to modulate CheA.

Mutant	WT	R391A	R386F	R386W	F395W
Purification	6mg	Precipitates	20mg	5mg	7mg
Oligomeric state(DSSO)	Trimer	X	Dimer	Monomer	-
Oligomeric state(MALS)	Trimer	X	Trimer, dimer	Monomer	-
CheA/W crosslinking	Yes	X	No	No	-
Tm CheA radioassays	Deactivating	X	Deactivating	No change	Deactivating

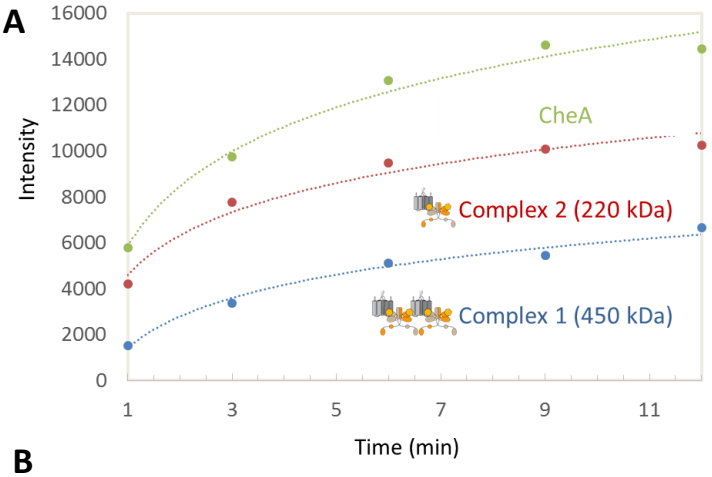
Fig 8: Results from mutational studies of the receptor foldons. In summary, mutations at R391 result in a mutant that is unstable and cannot be purified after *in vivo* expression. Mutations at R386 result in mutants that do not appropriately trimerize of make contacts with CheA and CheW. Mutation at F395W did not cause a change in CheA modulation.

2.2.5 Characterizing isolated ternary complexes

To determine the effects of ternary complex size (molecular weight) on CheA modulation, radioisotope assays were conducted with the two protein samples isolated by SEC. In comparison to samples of CheA alone, the 450 kDa and 220 kDa particle deactivates CheA 1.4-fold and 2.3-fold, respectively (Fig. 9). The difference in the effects on CheA modulation of the two particles is likely due to difference of the average number of receptor foldons associated with CheA. For the 450 kDa particle, the CheA dimers are on average associated with 1.5 receptor foldons while the 220 kDa particle has one foldon associated.

Low-resolution envelopes of the ternary complexes were generated via SEC-SAXS. As a control, samples of CheA and CheW were also used to ensure that the Porod volume and radius of gyration increase when receptor foldon is present. As expected, the Porod volume, Rg and estimated molecular weight increase when CheA and CheW are coupled to receptor (Fig 10). Furthermore, the presence of the P1 and P2 domains of CheA increase the volume of the envelope. When a hypothetical model of the ternary complex is rigid-body fit into the SAXS envelopes of the complex, it fits the expected size and shape.

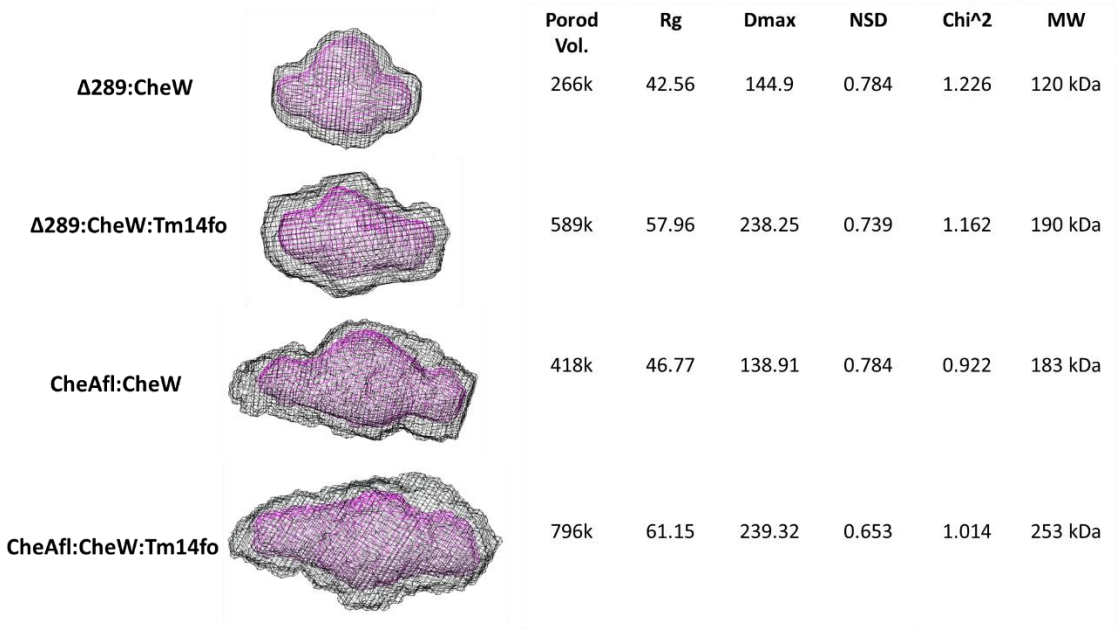
A



B

Sample	Relative [CheA-P]	[CheA-P] 95% confidence bounds	Rate constant (k min^{-1})	k min^{-1} 95% confidence bounds	[CheA-P]*k
Complex 1	23.4	+/- 14.22	0.181	+/- 0.352	**4.23**
Complex 2	30.63	+/- 81.1	0.139	+/- 0.751	**4.25**

Figure 9: (A) Radioisotope experiments that monitor CheA autophosphorylation demonstrate that the isolated ternary complexes differ to the extent that CheA activity is modulated. Specifically, the larger complex deactivates Tm CheA more than the smaller one. (B) Rate constants calculated from radioisotope experiments.



	Porod Vol.	Rg	Dmax	NSD	Chi^2	MW
Δ289:CheW	266k	42.56	144.9	0.784	1.226	120 kDa
Δ289:CheW:Tm14fo	589k	57.96	238.25	0.739	1.162	190 kDa
CheAfl:CheW	418k	46.77	138.91	0.784	0.922	183 kDa
CheAfl:CheW:Tm14fo	796k	61.15	239.32	0.653	1.014	253 kDa

Figure 10: SAXS experiments demonstrate that the size and shape of SAXS envelopes changes when CheA and CheW are reconstituted into ternary complexes with receptor foldons. The NSD values calculated for each envelope indicate that the samples are monodispersed. Furthermore, the molecular weight estimate calculated from the porod volume matches calculations from MALS experiments.

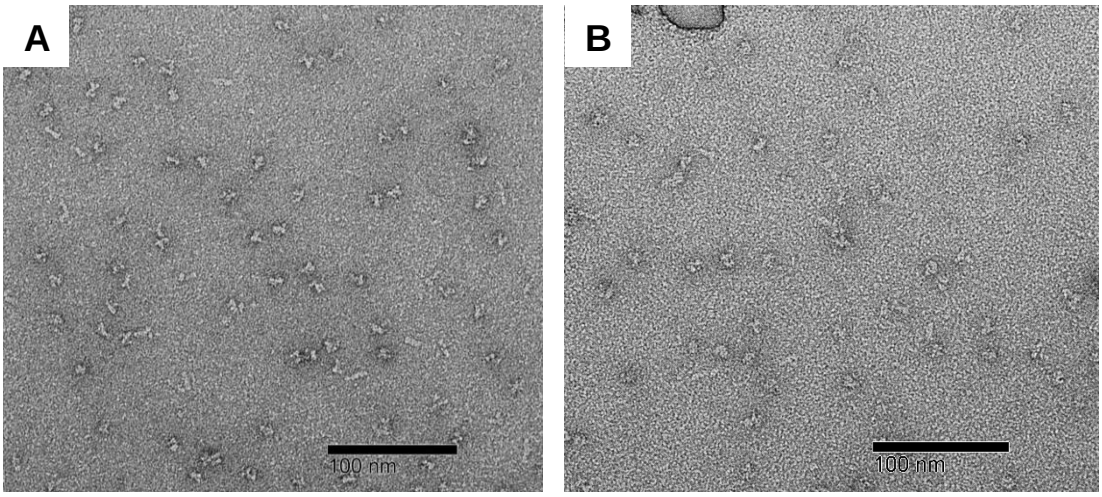


Figure 11: Negative stain electron microscopy experiments of the protein complexes indicate that (A) complexes are monodispersed. (B) When the complexes are allowed to incubate at RT overnight, the particles are not monodispersed and aggregates are present.

Electron microscopy (EM) micrographs of the 220 kDa ternary complex were examined to determine the homogeneity of the isolated complex. Inspection of the micrographs reveals that the particles are monodispersed in size and have a distinct U-shape (Fig. 11). Two-dimensional classification was attempted with this sample but was not successful due to relative heterogeneity of the particles.

2.2.6 Crosslinking ternary complexes

To increase the homogeneity of the isolated ternary complexes, the chemical crosslinker DSSO was used to stabilize protein:protein interactions. DSSO specifically cross-links the primary amine of lysine residues that are within 10 Å of each other in solution. Protein DSSO cross-linking followed by SDS-page reveals a high molecular-weight species is formed when a receptor:CheA:CheW complex is present, and that the ternary species is larger than a complex consisting of just CheA and CheW (Fig. 12).

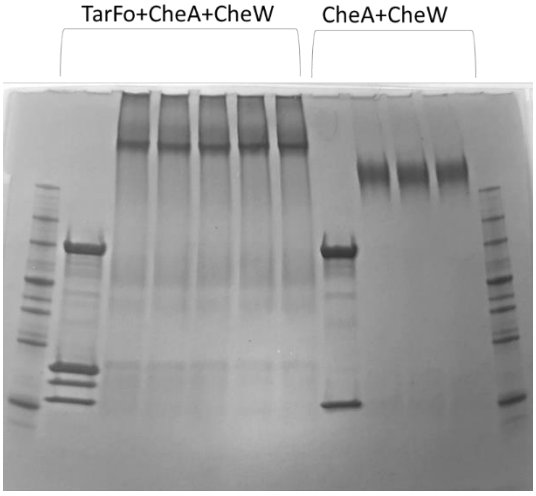


Figure 12: DSSO cross-linking followed by SDS-page demonstrates that receptor foldons mixed with CheA and CheW in a 1:1:1 ratio form a ternary complex with a distinct dominant species. Furthermore, there is an obvious molecular weight shift when receptor is added to a sample that only contains CheA and CheW.

MALS of crosslinked complex

The cross-linked ternary complex was isolated by SEC and the sample was further analyzed by SEC-MALS, SAXS, and EM. Analysis by SEC-MALS reveals that the isolated ternary complex elutes as a single peak and has an estimated molecular weight of 330 kDa, which corresponds

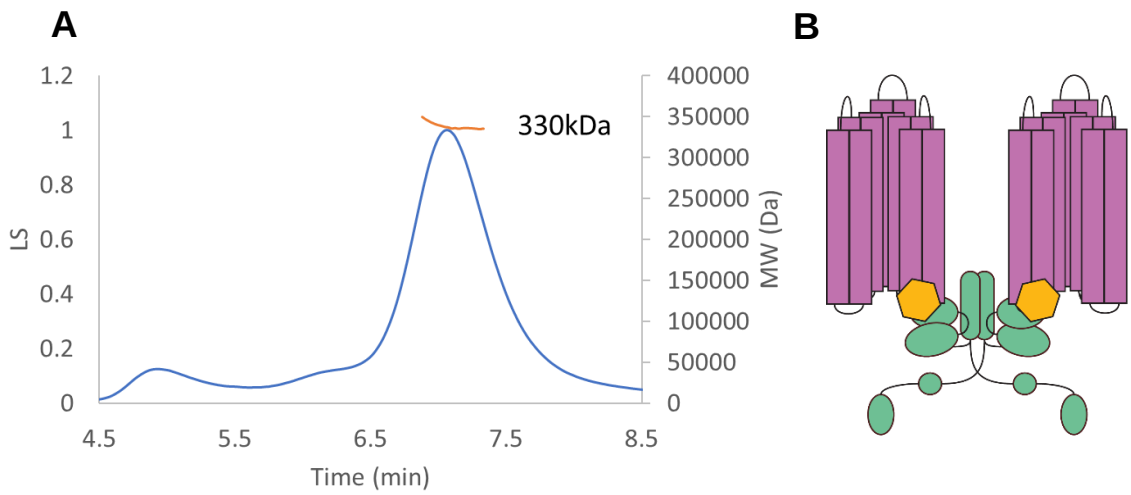


Figure 13: (A) MALS experiments of isolated crosslinked complexes demonstrate that the dominant cross-linked species is around 330 kDa, which corresponds to the 'core' chemotaxis complex consisting of 1 CheA dimer, 2 CheW subunits and 2 receptor foldons (B). Furthermore, the molecular weight of the complex is consistent over the entire SEC peak, indicating that the complex is homogenous.

to the molecular weight of the core chemotaxis complex consisting of one CheA dimer, 2 CheW proteins and two timerized foldons (Fig. 13). Furthermore, the molecular weight is stabilized over the entire peak indicating that the sample is compositionally homogenous. These results suggest that the 320 kDa core complex is stabilized by DSSO cross-linking and can then be isolated in high purity.

SEC-SAXS of cross-linked ternary complex

Molecular envelopes of the cross-linked ternary complexes were generated via SEC-SAXS experiments. Like the SEC trace from MALS experiments, the SEC trace of the 330 kDa complex reveals a single peak that is monodispersed (Fig. 14A). Rigid body fitting of the hypothetical structure of the ternary complex into the SAXS envelope validates that the particles fit the expected size and shape of the core complex (Fig. 14B).

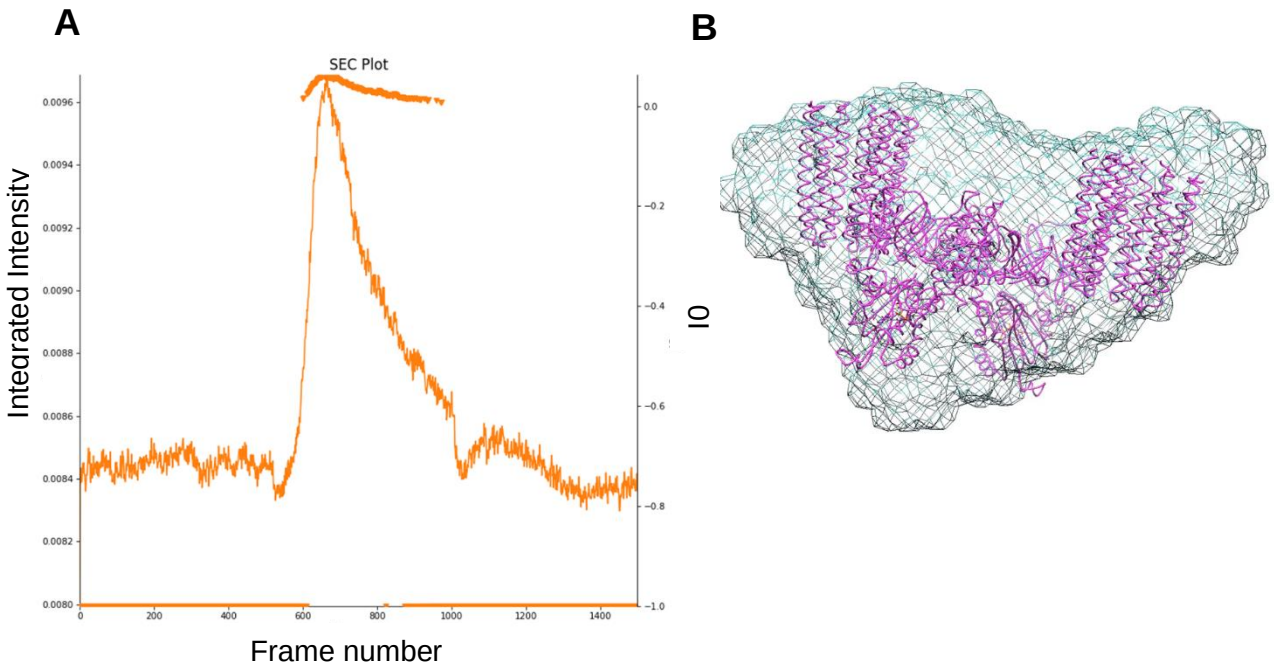


Figure 14: (A) SEC-SAXS elution profiles demonstrate that the isolated crosslinked complexes are homogenous and stable. (B) SAXS envelopes generated from scattering data of the cross-linked complexes fits the expected size and shape of the ternary core complex. The ternary complex was generated from crystal structure of CheA and CheW, and cytoplasmic fragments of receptors.

EM images of the crosslinked complexes

The homogeneity of the crosslinked complexes was further examined via cryo- electron microscopy (cryo- EM) experiments. Indeed, the particles in the micrographs were

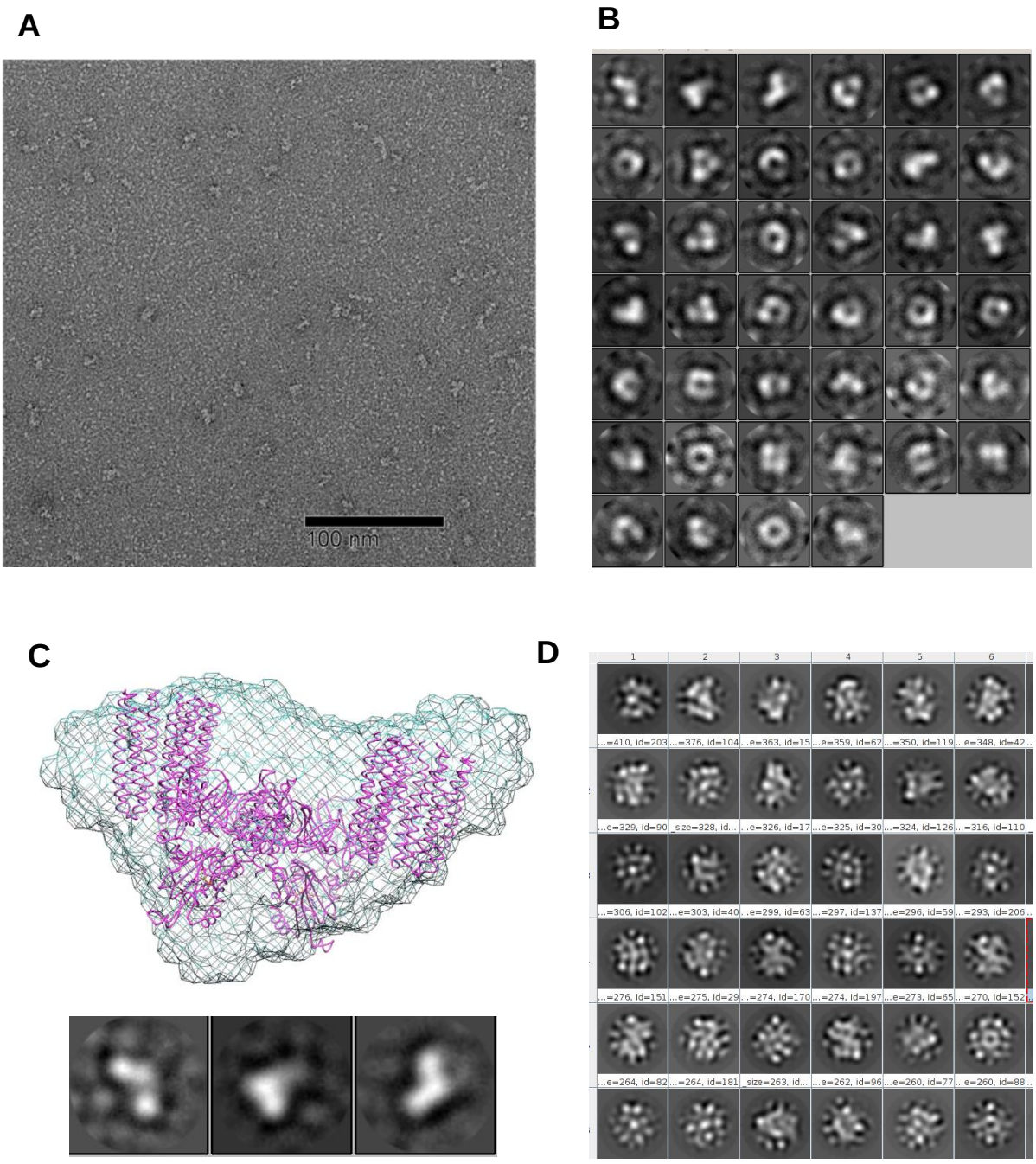


Figure 15: (A) Cryo-electron microscopy experiments of crosslinked complexes reveals particles that are monodispersed. (B) Analyses of the micrographs resulted in forty two-dimensional classifications of the particles. (C) The size and shape of the particles matches the SAXS envelopes and hypothetical structure of the ternary complex. (D) Two-dimensional classification of isolated complexes without chemical crosslinking was unsuccessful.

monodispersed and fit the expected size (Fig. 15A). Furthermore, 2D classification of these

particles results in classes that resemble the structure of the SAXS envelope (Fig. 15). Further optimization of the cryo- EM conditions are being conducted to attempt three-dimensional reconstructions of the particles.

2.2.7 Mass spectrometry of crosslinked complexes

Free CheA dimer

The small-molecule DSSO is a unique cross-linker in that it can be symmetrically cleaved when used in mass spectrometry electron-capture dissociation experiments. Analysis of the crosslinked fragments by MS/MS/MS after protease digestion identifies lysine residues that are crosslinked by DSSO and therefore come within ~20 Å of each other in solution (measuring from the lysine β-carbons). To investigate the structural changes that occur in CheA between the free kinase and the deactivated state, DSSO cross-linking followed by mass spectrometry was conducted on samples consisting of the free kinase and the isolated ternary complex (Tar foldon, CheA, CheW). Previous data suggests that in the free kinase, the P1 and P2 domains are highly flexible and are generally disassociated from the kinase core (P3-P5) while the deactivated state induces interactions among the P1/P2 domains with kinase core (Fig. 16).

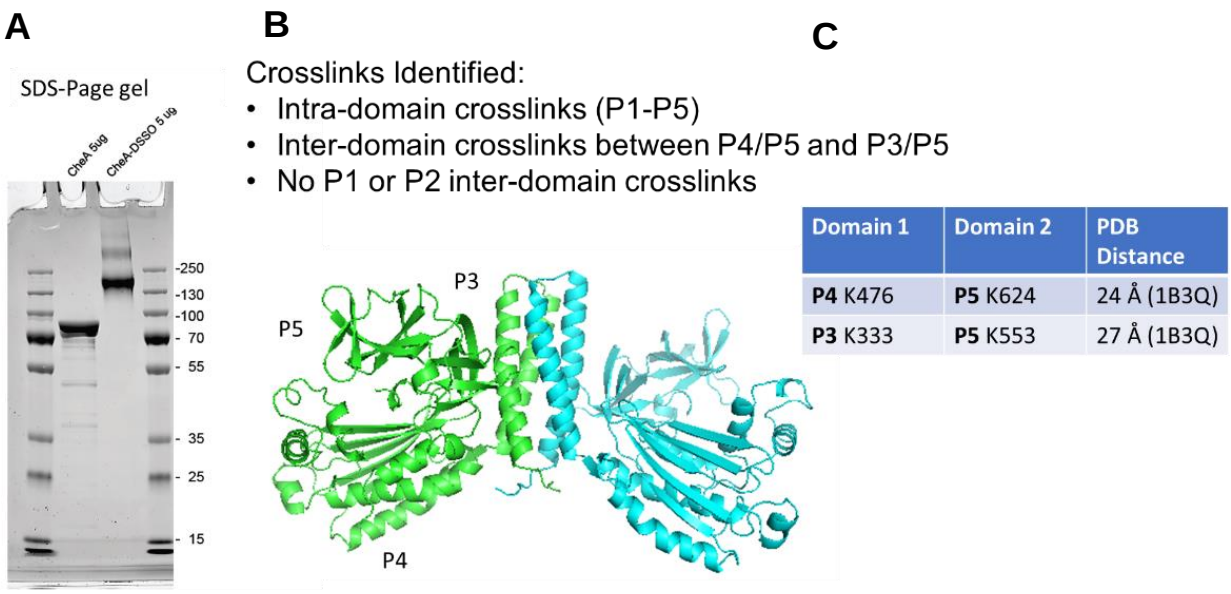


Figure 16: (A) DSSO cross-linking of CheA followed by SDS-PAGE reveals the formation of a 150 kDa dimer. (B) Mass spectrometry experiments of the crosslinked CheA dimer reveals inter-domain interactions among the kinase core (P3P4P5) but there are no inter-domain interactions identified with the P1 and P2 domains. (C) Based on published crystal structures, the inter-domain contacts identified by mass spec are within the expected distance constraints (<30 Å).

The DSSO cross-linking experiments of the free kinase identified cross-links that occurred via intra-domain lysines that are close to each other in available atomic protein structures (1B3Q). However, there were also inter-domain cross-links among the P4/P5 domains and P3/P5 domains. While many of these contacts are within the 30 Å range of the cross-linker according to atomic structures, several of the cross-links involve residues that are further than 30 Å away and remark on the high flexibility of the P4 and P5 domains in the free kinase (Appendix 1). In agreement with the experiments in Greenswag et al., no cross-links were identified that suggests interactions of the P1 and P2 domains with the kinase core (P3P4P5).

Ternary complex of Tar foldon with CheA and CheW

To determine the structural changes that occur from receptor deactivation, identical cross-linking/MS experiments were conducted with the ternary TarFo:CheA:CheW complexes. The cross-linking results of the ternary complex not only confirm protein:protein interactions present in protein crystal structures, but indicate the presence of interactions between the P1/P2 domains with the kinase core. Specifically, P1 and P2 crosslink to the P4 domain near the nucleotide-binding region. Interestingly, the lysine residues in the P1/P2 domains that crosslink to the P4 domain often cross-link with each other as well, suggesting that these three domains are in close association with each other in the kinase-off state (Fig. 17). Additionally, the P3 to

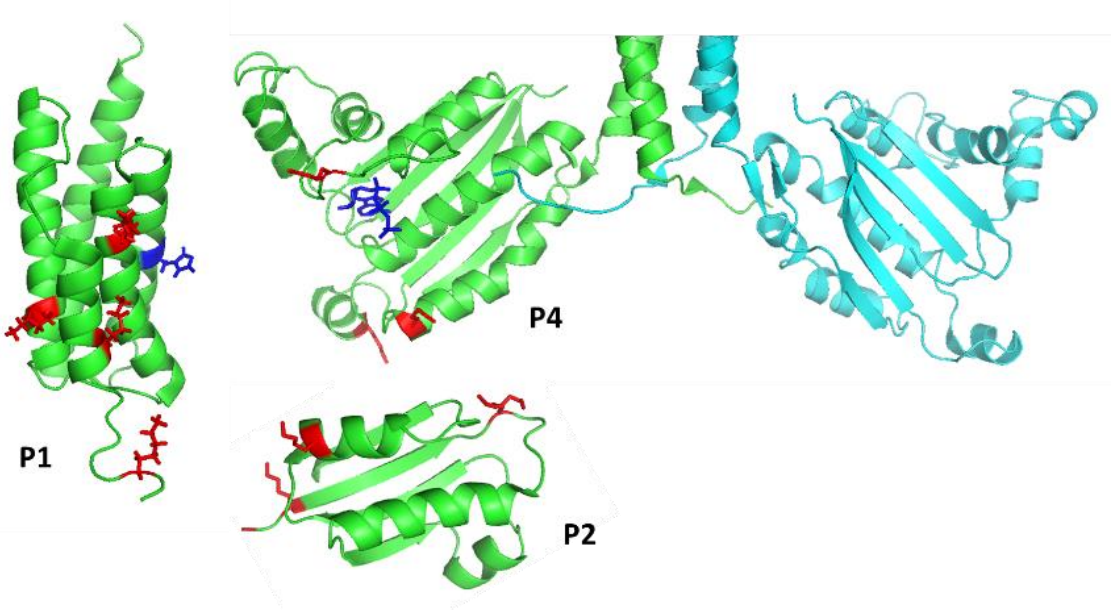


Figure 17: DSSO crosslinking of ternary complexes (Tar foldon:CheA:CheW) followed by mass spectrometry identifies inter-domain interactions among P1, P2 and P4. Since these interactions are not identified in the free kinase, they manifest from a conformational change that occurs in CheA when deactivated. Lysine residues in P1, P2 and P4 that interact with each other are highlighted in red.

P5 interactions present in the free kinase are no longer identified in the complex, which supports the hypothesis that the P5 domains are less flexible when bound to receptor (Appendix 1).

2.2.8 ESR DEER experiments with nitroxide-labeled CheA

P1 distant constraints

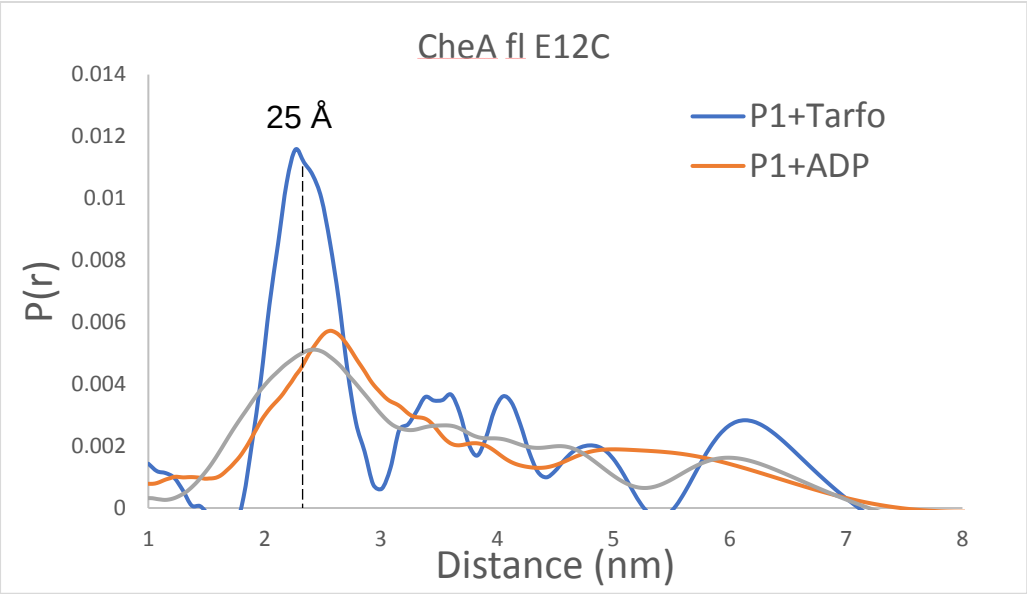


Fig 18: Distance distributions generated from ESR DEER experiments of CheA labeled with a nitroxide at position E12C demonstrate that a short distance of ~25 Å is present, which is indicative of an interaction between the P1 domains of the CheA dimer. Furthermore, this P1:P1 interaction is influenced by the presence of the deactivating Tar foldon.

Although the DSSO cross-linking/mass spectrometry experiments identified differences in inter-domain contacts of CheA, these crosslinked lysine residues can be within a 20 Å radius from one another in solution (measured from β carbons). Therefore, there are not sufficient restraints to accurately place the P1,P2,P4 domains in relation to one another in the ternary complex. To provide distance restraints for modeling the ternary complex, Double Electron Electron Resonance spectroscopy (DEER) experiments were conducted with nitroxide labeled CheA. Previous research with nitroxide-labeled CheA labeled on the P1 domain indicates the presence of unexpectedly short, ordered distances between the P1 domains in the CheA dimer. When CheA is labeled at residue S76C, short distances of ~25 Å are apparent and the distance distributions are shifted when CheA is in a complex with the Tar foldon and CheW (Fig. 18). Previous studies from Greenswag et al. have demonstrated the potential for a P1:P1 interaction

that is responsible for regulating CheA autophosphorylation. In these experiments, removal of one of the P1 domains from the CheA dimer significantly increases autophosphorylation, suggesting that the P1:P1 interaction is inhibitory and may reflect a deactivated conformation. Likewise, the 25 Å DEER peak becomes more dominant when CheA is in a deactivating complex, suggesting the interaction between the P1 domains is more prevalent in the deactivated state (Fig. 18).

P4 distant constraints

DEER distances between P4 domains in the CheA dimer have previously been reported but the distance distributions were not influenced by the presence of the receptor dimers used in these experiments[7]. In comparison, when CheA E387C-MTSL was measured in the presence of the Tar receptor foldon, a change in distance distribution from free CheA is observed. To determine the change in position of the nucleotide region of the P4 domain, DEER measurements were measured using a nitroxide-labeled ADP nucleotide, ADP-β-S-SL (see Chapter 4 for nucleotide synthesis and characterization), reconstituted into the full-length kinase. When the ADP-β-S-SL is reconstituted into the free kinase, either full length or ΔP1P2, no clear distances can be measured. This effect is likely due to the innate flexibility of the P4 domains. However, when the kinase is in the ternary complex distinct distances are apparent, demonstrating that P4 is less dynamic when CheA is in a deactivating complex (Fig. 19). Furthermore, the distributions are influenced by the presence of the P1 and P2 domains, indicating that P1/P2 interacts with P4 in the deactivated state and influences the distance between the nucleotide binding sites. Specifically, when the P1 and P2 domains are not present, a short distance of about ~35 Å is dominant. However, when the full-length kinase is measured, a distance of ~60 Å is present

(Fig. 19). Therefore, it's likely that the P1/P2 domains anchor the P4 domains further away from each other and closer to the P5 domain.

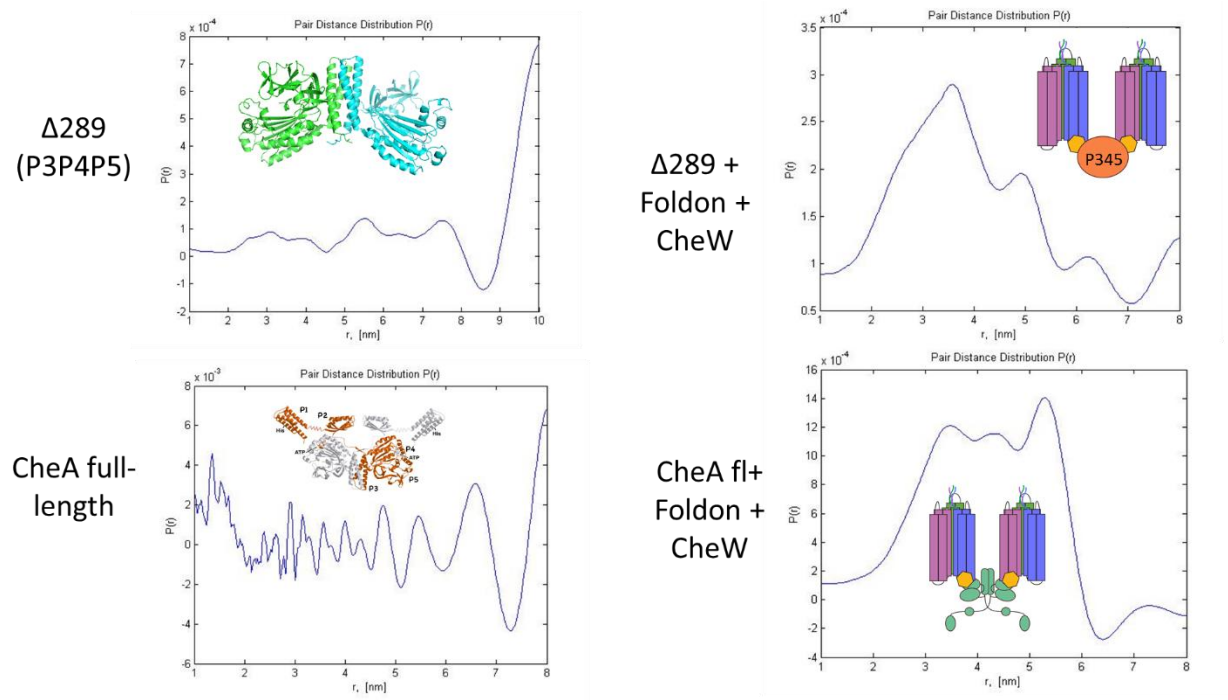


Figure 19: ESR experiments of CheA reconstituted with the ADP analog ADP-NO in the nucleotide binding pocket demonstrate that the P4 domain is less mobile in the presence of deactivating receptor. Furthermore, the presence of the P1 and P2 domains induce longer distances in the distribution.

P4 to P5 distant constraints

To further explore the position of the P4 domain in the ternary complex, the P5 domain was labeled on residue E588C and ADP-β-S-SL was reconstituted into the P4 domain. In the free kinase, a distance of ~62 Å is measured. However, when CheA is present in a ternary complex, the distance distribution shows a ~25 Å distance (Fig. 20). Again, these data suggest that the P4 domains move closer to the P5 domain in the kinase-off conformation.

2.2.9 Structure of P4 domain with a synthetic ADP analog

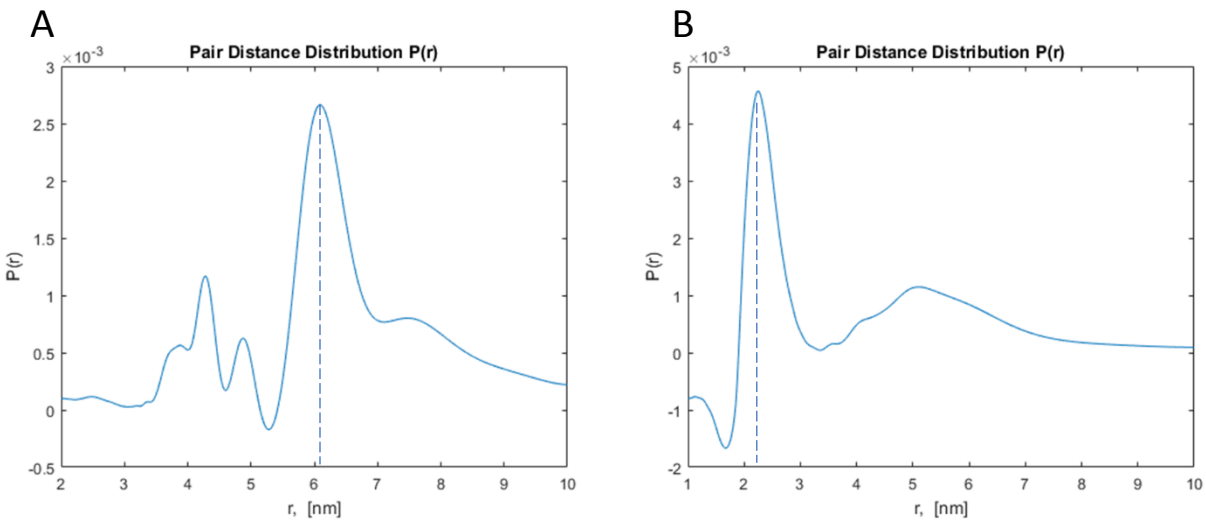


Figure 20: ESR DEER experiments with (A) free TmA E588C-MTSL with reconstituted ADP-β-S-SL and (B) in a complex with deactivating receptor.

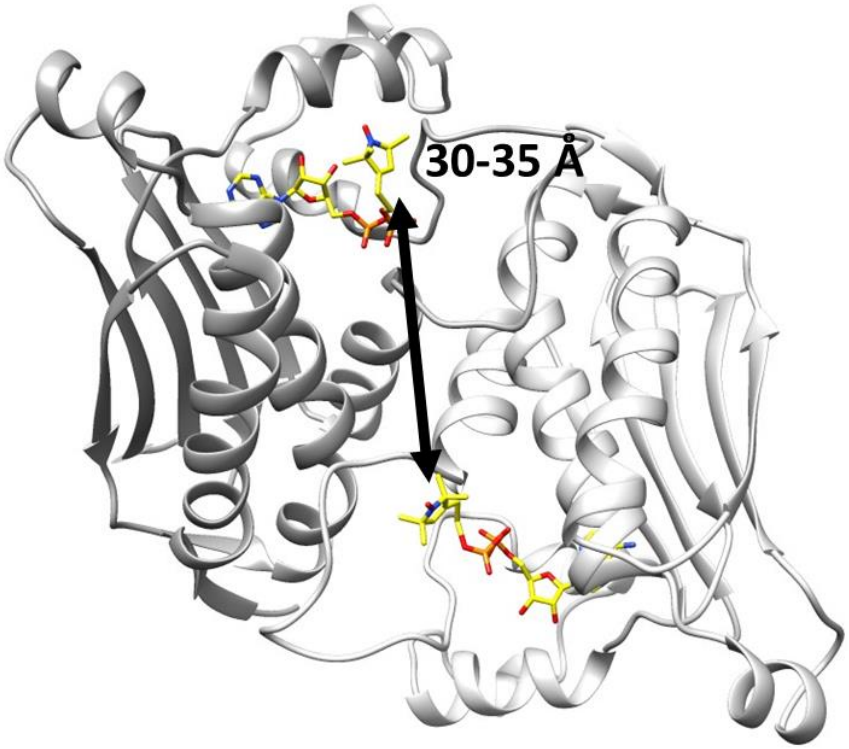


Figure 21: The crystal structure of *T. maritima* P4 domain reconstituted with ADP-NO. The P-loops are pulled away from the nucleotide pocket to stabilize the dimer interaction. The distance between the synthetic nucleotides is between 30-35 Å, accounting for nitroxide flexibility in the molecule.

The P4 domain of CheA was co-crystallized with ADP-β-S-SL using the crystallization

conditions previously reported in Bilwes et al[8]. The structure of CheA[ADP-β-S-MTSL] reveals the P4 homo-dimer reported in Bilwes et al. that was produced by co-crystallization with the ATP analog TNP-ATP (Fig. 21)[8]. Interestingly, the P-loop of the P4 domain interacts with residues in the adjacent subunit so that the loop is pulled away from the nucleotide-binding site. Furthermore, the residues at the P4 dimer interface are highly conserved, suggesting that the formation of this dimer may be evolutionarily preserved. In support of the DEER measurements in Figure 19, the distance between the nucleotide sites is approximately 30-35 Å.

2.2.10 Cysteine cross-linking of P4 mutants

Cysteine cross-linking assays were conducted with full-length CheA to determine if the P4 dimer present in the CheA[ADP-β-S-MTSL] crystal structure is present *in vitro*, and if this interaction is regulated by receptor foldons. Residues V509 and N112 were mutated to Cysteine based on their location at the dimer interface—both Cysteine residues could crosslink to the same residue on the adjacent subunit. However, in previous crystal structures (1B3Q) the V509 and N112 are about 60 Å from each other. For both sites, the relative amount of cross-linked dimer significantly increases in the presence of deactivating receptor foldon (Fig. 22). Specifically, there is about 2-fold more crosslinking between the residues when CheA is in a deactivating complex.

2.2.11 Two-dimensional crystallization of the ternary chemotaxis complex

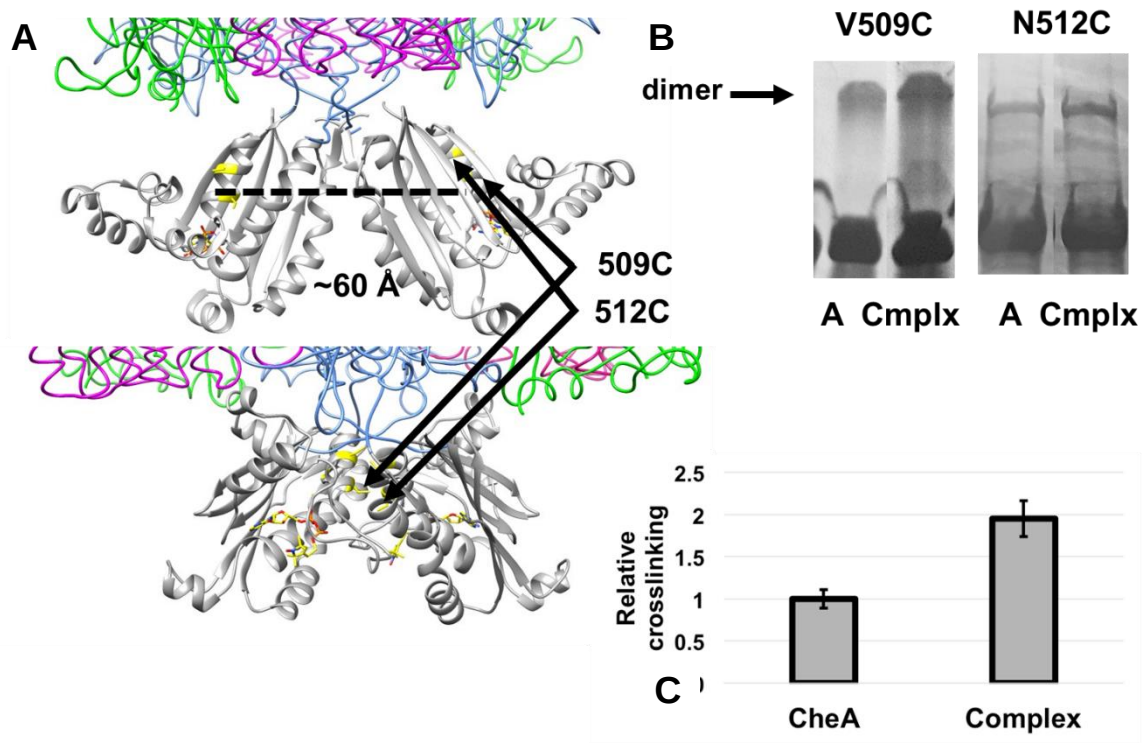


Figure 22: Cysteine cross-linking experiments with residues at the P4 dimer interface seen in crystal structures indicate that this dimer interaction is regulated by deactivating foldon. (A) In available crystal structures V509 and N512 are around 60 Å apart (top image). In the new dimer structure, they are 5 Å apart (bottom image). (B) Cross-linking followed by SDS-PAGE gel analysis reveals a higher abundance of CheA dimer when deactivating receptor is present. (C) Quantification of the gel intensities from A indicate a significant difference in cross-linking.

Attempts to crystallize the ternary complex consisting of receptor foldon:CheA:CheW through traditional three-dimensional crystallography methods were unsuccessful. Therefore, a newly developed method for forming two-dimensional crystals was used. This method, called lipid monolayer crystallization, induces crystal packing of particles in solution through electrostatic interactions with a lipid monolayer at the air-water interface of the sample droplet[9]. Interaction between the protein particles and lipids are induced by incorporating a synthetic lipid with a nickel-containing head-group in the monolayer and expressing the proteins with 6X-PolyHis tag.

Formation of two-dimensional crystal is tested by negative stain electron microscopy. Like three-dimensional methods, the buffer conditions of the solution are optimized to encourage ordered crystal growth.

Lipid monolayer crystallization has been used to determine the structure of a ternary receptor:CheA:CheW complex but the resolution is limited to >11 Å[10]. However, this limit in resolution may be attributed to the fact that receptor dimers were used to produce the complexes, and these dimers consists of longer portions of the receptor than the engineered foldons. Using the foldons not only stabilizes the receptor trimer-of-dimers state present *in vivo* but they consist of only the CheA/CheW binding domain, making them less flexible.

Although initial attempts at 2D crystallization resulted in amorphous aggregates of protein on the monolayer, further optimization results in slightly ordered protein arrangements. Specifically, the optimized electron micrographs show ring-like structures that resemble protein formation present in *in vivo* arrays. Furthermore, Himes et al. found similarly patterned micrographs in early attempts to crystallize the ternary complex with cytoplasmic fragments of receptor dimers[10]. Therefore, further optimization of the buffer conditions can be tested to form two-dimensional crystals suitable for structural determination.

2.3 Discussion

Attempts to structurally analyze the core chemotaxis complex has been limited due to the insolubility and complex oligomeric state of transmembrane receptors[2]. Here we have engineered receptor mimetics that circumvent these issues by incorporating soluble fragments of receptors into protein variants that pre-form a trimer-of-dimer arrangement[3]. Importantly, these mimetics, referred to as receptor foldons, strongly interact with and modulate the behavior of the histidine kinase CheA. Furthermore, ternary complexes of the foldons with CheA and CheW can be isolated in milligram quantities and are suitable for traditional structural

experiments. By mutating residues in the foldon that are known to participate in protein:protein interactions with CheA/CheW, we were able to demonstrate the integrity of the foldons is compromised in a manner predicted by published *in vivo* experiments.

Further analysis of the isolated ternary foldon:CheA:CheW complexes by MALS and SEC-SAXS reveals that they are relatively stable and monodispersed. Indeed, micrographs of the complexes generated via negative stain electron microscopy reveal particles that are monodispersed in size with a noticeable butterfly-shaped structure. However, two-dimensional reconstructions of these particles were unsuccessful due to insufficient homogeneity of the particles.

Addition of the chemical crosslinker DSSO to the complexes stabilizes the 'core' unit of the chemoreceptor array, which consists of a CheA dimer with two CheWs and two receptor foldons. SEC-MALS and SEC-SAXS experiments of the isolated crosslinked complexes demonstrates that these complexes are extremely homogenous and stable. Furthermore, SAXS envelopes of the particles fit well with the expected size and shape of the complex. Cryo-electron microscopy experiments of the crosslinked complexes successfully generated two-dimensional classification of the particles. Comparison of these classifications with the SAXS envelopes reveals a similar size and shape of the particles. Further optimization of sample buffer and freezing conditions are currently being done to increase the contrast of the cryo-EM micrographs to conduct three-dimensional reconstruction experiments of the particles.

Mass spectrometry of DSSO cross-linked CheA and ternary complexes reveals key conformational changes that occur in CheA when it's locked in a kinase-off state. When CheA is present as a free protein, the P1 and P2 domains do not participate in any inter-domain interactions. However, when CheA is coupled to a deactivating foldon, the P1 and P2 domains not only interact with each other but they also bind the P4 domain near the nucleotide binding pocket. Interestingly, the same inter-domain lysine residues in the P1 and P2 domains that

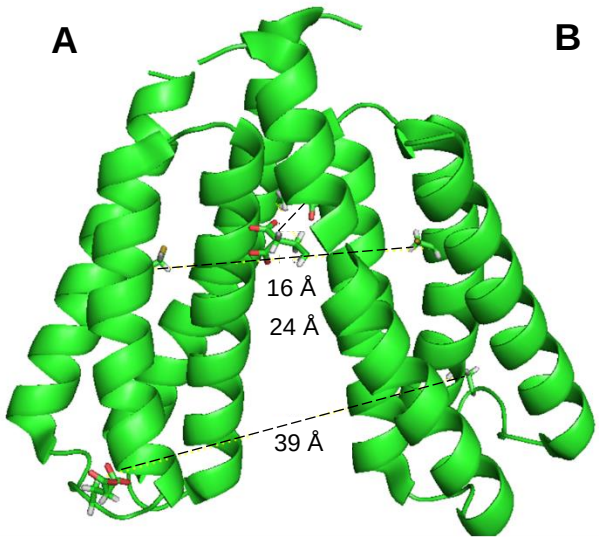
cross-link with one another also interact with the P4 domain, suggesting that the P1 and P2 domains may form a structure around the nucleotide pocket and preclude it from binding ATP. Furthermore, this data demonstrates that the P1 domain is anchored to the kinase core in the deactivated state, which may prevent the substrate histidine from accessing the nucleotide pocket and being phosphorylated. Indeed, both scenarios would provide a potential mechanism for CheA deactivation due to induced domain rearrangements.

The CheA domain rearrangements determined by the mass spectrometry experiments concur with the density maps of whole cell tomograms. The maps demonstrate that in the deactivated state, there is an increase in electron density below the P5-CheW rings that corresponds to the placement of the P1, P2 and P4 domains[11]. Likewise, interactions among these domains are only seen in the mass spectrometry data when CheA is coupled to a deactivating receptor.

Although the DSSO cross-linking followed by mass spectrometry experiments have defined differences in inter-domain interactions in CheA, the distance restraints supplied by the cross-links are not sufficient for producing an accurate model of the ternary complex. This is due to the fact that the DSSO crosslinker has a long spacer arm. To develop a model of the ternary complex with additional restraints, double electron-electron resonance experiments were conducted with nitroxide label CheA. When the P1 domain was labeled on E12C, distances between 20-30 Å were dominant. This result was unexpected since the P1 domains are located on the end of long, highly flexible linkers and therefore are expected to be far from one another if the domains are moving freely. Furthermore, the presence of deactivating receptor causes the distance around 22 Å to narrow significantly. The regulation of the 20-30 Å distance by receptors suggests that CheA deactivation involves an interaction between P1 domains in the CheA dimer. Additionally, when the E12C distributions are compared to previous DEER data with P1 labeled on two different residues (C63, E76C), they also both produce distances that

are unexpectedly short. Interestingly, when these distances are compared to inter-domain distances in a P1 crystallographic dimer (1TQG), they are in relative agreement with one another (Fig. 23). Therefore, it is possible that this crystallographic dimer is representative of intra-dimer contacts that occur in CheA *in vitro*.

The presence of a P1:P1 interaction in the CheA dimer is supported by *in vitro* and *in vivo* experiments that demonstrate an increase in CheA activity when one of the P1 domains is removed[12]. Since the ESR DEER data suggest that the presence of a P1 dimer is increased by deactivating receptors, it's likely that this inhibitory interaction is induced by receptors as a method for modulating CheA activity. Furthermore, this substrate histidine on P1 (H45) is proximal to the dimer interface, which suggests that the dimer interaction may inhibit activity by preventing His45 from interacting with ATP at the P4 domain.



B

Labeled site	EPR distance	Dimer structure distance
E12C	24 Å	23.3 Å
E63C	30 Å	31.4 Å
E76C	62 Å	45.7 Å

Measured from alpha carbon +7Å for MTSL labels

Figure 23: Summary of ESR results from P1 labels. (A) A crystallographic dimer of P1 is present in PDB 1TQG. (B) The experimentally determined ESR distances match the distances present in the P1 dimer.

Although there are several crystal structures of the core CheA domains P3-P5 (1B3Q, 2CH4, 3UR1, 4JPB, 4XIV), the exact position of the P4 domain changes in relation to the P5 domain due to its dynamic nature. In tomography maps of *in vivo* arrays, the density corresponding to the P4 is sparse due to delocalization of this domain across the array[13]. It's hypothesized that the mechanisms for CheA on-off switching involves rearrangement of this domain, specifically near the nucleotide binding pocket. To better understand the conformational changes that may occur in the P4 domains between activity states, we utilized a synthetic nucleotide that contains a nitroxide spin label to directly label the ligand-binding site for DEER experiments without inducing functional perturbations through site-directed mutagenesis (see chapter 4 for a detailed description of this small molecule). When the synthetic nucleotide, referred to as ADP-β-S-SL, is reconstituted into the free CheA dimer no apparent distance can be seen in the DEER spectra. However, when CheA is in complex with the deactivating receptor foldon, distances are apparent. Furthermore, the distance distributions change when the P1 and P2 domains are removed. Specifically, longer distances of ~60 Å are present when the kinase is intact (P1-P5) but only shorter distances ~35 Å are present when P1 and P2 are removed. This data further supports a model where the P1 and P2 domains interact with P4 in the deactivated state to stabilize its position in the complex.

The ~35 Å distance between the nucleotide sites seen in the DEER data is further supported by a crystal structure of the isolated P4 domain that reveals a symmetric P4 dimer. The distance between the nucleotide pockets in this structure are ~ 37 Å, which is in agreement with the DEER measurements. In this dimer, the P-loop of the P4 domain is pulled away from the nucleotide pocket through interactions of highly conserved residues.

To test whether this P4 dimer occurs *in vivo,* residues were selected for cysteine-crosslinking experiments based on location and relative proximity at the P4 dimerization site. Two sites were chosen for cross-linking experiments, V509C and N112C, as these residues are

located at the dimer interface and are ~5 Å away from each other. Indeed, cross-links corresponding to the CheA dimer using both sites are more abundant in the presence of receptor, supporting the hypothesis that this P4:P4 interaction plays a regulatory role in CheA modulation. Specifically, this dimer is representative of a CheA conformation in the deactivated state. The presence of this dimer could explain the negative allosteric binding of ATP to the P4 nucleotide sites in the CheA dimer—binding of nucleotide in one of the sites could perturb the second site since they are immediately adjacent to one another. Furthermore, molecular dynamics simulations and disulfide cross-linking experiments suggest close association of the P4 domains in the CheA dimer[10,14].

Further DEER experiments were done to utilize the ADP-β-S-SL label and determine the placement of the P4 nucleotide pocket in relation to the P5 domain. To accomplish this, the P5 domain was labeled on residue E588C with an MTSL spin label, and ADP-β-S-SL was reconstituted into the P4 nucleotide pocket. In the absence of receptor, a distance of ~60 Å is apparent in the distribution. This distance may be present due to the fact that the P5 domains are flexible in the free kinase and a portion of them have closer to one another. Interestingly, when this sample forms a complex with deactivating foldon, the 60 Å distance is no longer present but there is a much short distance around ~25 Å. This data suggests that in the deactivated state, the P4 nucleotide pocket moves closer to the P5 domain. The fact that the DEER data indicates the P4 domains come closer to each other and to the P5 domain in the deactivated state initially seems contradictory. However, it is likely that the P4 domain samples these two different conformations in the deactivated state but the P4-P5 distance is more dominant in the double-labeled sample (P5 E588C, P4 ADP-β-S-SL) since negative cooperativity for ligand binding in the P4-P4 labeled sample (P4 ADP-β-S-SL) yields a significantly weaker signal.

In summary, the cross-linking/mass spectrometry and DEER experiments support a model where the CheA domains self-associate and form a much more rigid structure in the deactivated state than the free kinase. Specifically, these experiments demonstrate that the P1 and P2 domains associate with the kinase core in the deactivated state but not free kinase. Additionally, these results suggest that deactivating receptors induce P1:P1 and P4:P4 dimers that may play important regulatory roles in modulating CheA autophosphorylation and allosteric binding for nucleotide substrate.

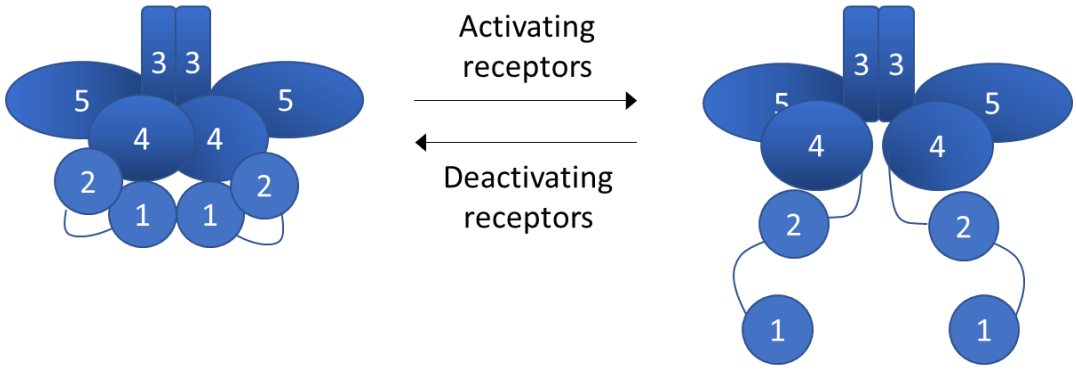


Figure 24: A model for CheA conformational changes in the kinase-off and kinase-on states. In the deactivated state the P4 and P1 domains form inter-subunit dimers and the P1, P2 and P4 domains interact with each other. In the activated state the P1, P2 and P4 domains are more mobile and interact with each other.

2.4 Methods and Materials

2.4.1 Protein purification

All proteins were cloned into pet28a vector containing a kanamycin resistance marker and transformed into *E. coli* (DE3) competent cells. After plating the transformation on agar containing kanamycin, a single colony was chosen for cell culture growth. 8 liters of cells were grown at 37°C until an O.D. at 600nm reached 0.6. 1mM IPTG was added to each flask at the cells were grown for 16 hours at room temperature. After the cells were pelleted via centrifugation, they were resuspended in 50ml of lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5mM Imidazole) and

sonicated for 6 minutes. The lysed cells were then centrifuged at high speed to remove the insoluble fraction from the lysate. The lysate was then run over a Ni-NTA affinity resin, washed with 20 mM Imidazole buffer to remove non-specific binding proteins, and finally eluted using buffer with 200 mM imidazole. The eluted protein was further purified by size-exclusion chromatography (prep-grade Sephadex s200). The fractions containing the protein of interested were pooled and concentrated via high-speed centrifugation.

2.4.2 Preparation and isolation of ternary complexes

Ternary complexes consisting of receptor foldon, CheA and CheW are isolated in milligram quantities by mixing the composite purified proteins in a 1:1:1 stoichiometric ratio in 50 mM Tris or Hepes pH 7.5, 250 mM KCl, 5mM $MgCl_2$, 10% glycerol and allowing the mixture to incubate at 4°C for at least 3 hours. After incubation, the sample is injected onto two prep-grade Sephadex columns (s200 and s300) attached to run in tandem with the elutant is collected in 6 ml fractions. Samples of the fractions were then taken to analyze the protein components in each fraction by SDS-PAGE. Fractions of interest were concentrated to about 8 mg/ml via high-speed centrifugation, flash frozen, and stored at -80°C for future experiments.

2.4.3 Cross-linking of ternary complexes

Crosslinking of ternary complexes to stabilize the formation of the core chemotaxis complex is accomplished by first incubating the receptor foldons with CheA and CheW in a 1:1:1 stoichiometric ratio to a final concentration of 10 µM of each protein in 50 mM Hepes pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 100 µM ADP, 10% glycerol. After incubation for at least 3 hours at 4°C, the chemical crosslinker DSSO is added to the protein mixture to a final concentration of 1 mM and crosslinking proceeds for 30 min to 1 hour at room temperature. The cross-linking reaction is quenched by the addition of Tris pH 8.0 to a final concentration of 20 mM. The reaction mixture is run through size-exclusion chromatography on two prep-grade sephadex columns (s200 and

s300) attached to run in tandem with a 6 ml fraction volume. Fractions of interest were selected to analyze the complex molecular weight and homogeneity by SDS-PAGE.

2.4.4 Characterization of complexes by MALS

Reverse-phase chromatography coupled to multi-angle light scattering experiments was used to determine the molecular weight of isolated ternary complexes. For each sample, 5mg/ml of protein was injected onto a Phenomenex reverse-phase column pre-equilibrated with 50mM Tris pH 7.5, 150 mM NaCl at room temperature. Before injection, each protein sample was buffer exchanged with the equilibration buffer to prevent contributions of buffer components from influencing the molecular weight calculations. Bovine serum albumin (sigma) was used as a protein standard. Wyatt technologies ASTRA 6 program was used for data analysis and molecular weight calculations.

2.4.5 Characterization of complexes by SAXS

Isolated ternary complexes were analyzed by size-exclusion chromotography (SEC) small-angle x-ray scattering (SAXS) to determine sample quality, particle flexibility in solution, and generate a low-resolution molecular envelope of the particles. Data were collected at CHESS G1 line on the Finger Lakes CCD detector. For each sample, ~3 mg/ml of protein was injected onto an Superdex 5/150 size-exclusion increase column pre-equilibrated with sample buffer (50 mM Tris pH 7.5, 150 mM NaCl) coupled to a SAXS beamline. The eluted sample was collected with 2 second exposure times to the x-ray beam at a flow rate of 0.1 ml/min. RAW56 and Primus57 were used to generate Kratky plots. Envelope reconstructions were calculated using ATSAS58 programs. Ten models were independently generated and then averaged into a consensus envelope. Hypothetical models of the ternary complex were fit into the envelopes via Chimera rigid body fitting algorithm.

2.4.6 Characterization of complexes by cryo-electron microscopy

Negative-stain electron microscopy experiments were conducted using samples of the isolated complexes to assess sample homogeneity and complex stability. Copper mesh grids (either 300 or 200) were glow discharged for 5 minutes at 5 mA to increase hydrophilicity. Protein sample at a concentration around 1 µg/ml was applied onto the grid for 45 seconds and blotted. The grid was washed by applying a droplet of water and then blotting. The grid was then stained with 0.5% Uranyl Formate solution for 1 minute and then blotted. The grids were imaged on a transmission electron microscope at 80 kV.

2.4.7 Mass spectrometry of crosslinked complexes

Chemical cross-linking with DSSO followed by mass spectrometry was used to examine CheA domain arrangements in the free kinase and ternary complex (Tar receptor foldon:CheA:CheW). After following the procedure for chemical cross-linking (in Section 2.2.3). The samples were run on an SDS-PAGE gel to separate the desired complex from the remaining composite proteins in the sample. The desired band was then cut out of the gel and further processed for MS/MS/MS analysis. The gel slice was washed with water/acetonitrile and air dried before being reduced and alkylated. A double protease digest using trypsin followed by chymotrypsin was used to increase the number of protein fragments that could be detected. The concentration of sample was calculated and then ran on LC MS/MS.

2.4.8 ESR DEER of labeled complexes

For PDS measurements, four pulse double electron electron resonance (DEER) experiments were conducted at 60 K on a 17.3 GHz FT EPR spectrometer, which is modified to perform PDS experiments[15]. The baseline used for data processing was approximated by a linear polynomial. A new singular value decomposition (SVD) method[16] is then applied to obtain distance distributions. Traditionally, Tikhonov regularization is used that solves for the complete distance distribution all at once. However, the regularization method is a compromise between a desired distance distribution and avoiding unstable solution, leading distributions with low resolution. The SVD method finds the optimal distribution value at each distance or distance range by

determining each of their different singular value cut-offs. The new method ensures optimal convergence at all distance ranges, while preventing a premature or unstable solution at some or all distance ranges. The method yields distance distributions without any spurious peaks.

2.4.9 Lipid monolayer crystallization of ternary complexes

To produce two-dimensional crystals of a receptor foldon:CheA:CheW complex, templating of the complexes on a lipid monolayer was attempted. A stock solution of isolated ternary complexes (~20 mg/ml) from SEC was diluted to ~0.75 mg/ml in candidate crystallization buffer solutions. 6 µl of the protein solution was placed in Teflon crystallization blocks and 1 µl of lipids containing a synthetic lipid with a nickel containing head group (DOGS-NTA) was placed on the surface of the droplet using a glass syringe. The amphipathic nature of phospholipids spontaneously creates a lipid monolayer on the surface of the droplet with the hydrophobic tails of the lipid extending away from the solution. Crystallization was allowed to occur overnight at 25°C in the dark. The lipid monolayer was transferred into electron microscopy grids by placing the hydrophobic surface of the grid on the lipid monolayer. The grids were stained with either 2% Tungstosilic acid or 1% Uranyl Acetate, and then washed with water to remove excess stain. The grids were air dried for 1 hour and then imaged on a Morgagni transmission electron microscope at 80 kV.

2.4.10 P4 crystallization

See Chapter 4 for the methods for P4 crystallization and refinement statistics.

2.5 References

1. Briegel, a. *et al.* Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. *Proc. Natl. Acad. Sci.* **109,** 3766–3771 (2012).

2. Boldog, T., Grimme, S., Li, M., Sligar, S. G. & Hazelbauer, G. L. Nanodiscs separate chemoreceptor oligomeric states and reveal their signaling properties. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11509–11514 (2006).

3. Greenswag, A. R. *et al.* Preformed Soluble Chemoreceptor Trimers That Mimic Cellular Assembly States and Activate CheA Autophosphorylation. *Biochemistry* **54,** 3454–68 (2015).

4. Mcnally, D. F. & Matsumura, P. Bacterial chemotaxis signaling complexes: Formation of a CheA/CheW complex enhances autophosphorylation and affinity for CheY (protein-protein interaction/protein phosphorylation/Escheichia coil). *Biochemistry* **88,** 6269–6273 (1991).

5. Mowery, P., Ostler, J. B. & Parkinson, J. S. Different signaling roles of two conserved residues in the cytoplasmic hairpin tip of Tsr, the Escherichia coli serine chemoreceptor. *J. Bacteriol.* **190,** 8065–8074 (2008).

6. Ortega, D. R. *et al.* A phenylalanine rotameric switch for signal-state control in bacterial chemoreceptors. *Nat. Commun.* **4,** 1–8 (2013).

7. Bhatnagar, J. *et al.* Structure of the ternary complex formed by a chemotaxis receptor signaling domain, the CheA histidine kinase, and the coupling protein CheW as determined by pulsed dipolar ESR spectroscopy. *Biochemistry* **49,** 3824–41 (2010).

8. Bilwes, A. M., Quezada, C. M., Croal, L. R., Crane, B. R. & Simon, M. I. Nucleotide

binding by the histidine kinase CheA. *Nat. Struct. Biol.* **8,** 353–360 (2001).

9. Yeager, M., Dryden, K. A. & Ganser-pornillos, B. K. Lipid Monolayer and Sparse Matrix Screening for Growing Two-Dimensional Crystals for Electron Crystallography: Methods and Examples. **955,** 527–537 (2013).

10. Cassidy, C. K. *et al.* CryoEM and computer simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. *Elife* **4,** 1–20 (2015).

11. Briegel, A. *et al.* NIH Public Access. **14,** 384–399 (2010).

12. Greenswag, A., Muok, A., Li, X. & Crane, B. HHS Public Access. **427,** 87–92 (2016).

13. Briegel, A. *et al.* Structure of bacterial cytoplasmic chemoreceptor arrays and implications for chemotactic signaling. 1–16 (2014). doi:10.7554/eLife.02151

14. Careaga, C. L. & Falke, J. J. Structure and dynamics of Escherichia coli chemosensory receptors. Engineered sulfhydryl studies. *Biophys. J.* **62,** 209-16; discussion 217–9 (1992).

15. Borbat, P. P., Crepeau, R. H. & Freed, J. H. Multifrequency Two-Dimensional Fourier Transform ESR : An X / Ku – Band Spectrometer. **167,** 155–167 (1997).

16. Srivastava, M. & Freed, J. H. Singular Value Decomposition Method to Determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance. 5648–5655 (2017). doi:10.1021/acs.jpclett.7b02379

3

Chapter 3

A soluble di-iron protein functions as an iron and oxygen sensor for bacterial chemotaxis

3.1 Introduction

During bacterial chemotaxis methyl-accepting chemotaxis proteins (MCPs), or chemoreceptors, convey changes in the external chemical environment to a cytoplasmic phosphorelay system that ultimately directs rotation sense of the flagellar motor. Although canonical chemoreceptors contain transmembrane regions and periplasmic ligand-binding domains, several receptor classes found in, for example, *Thermotoga maritima*, *Pseudomonas aeruginosa*, *Rhodobacter sphearoides*, *Bacillus subtilis,* and *Treponema denticola* do not[1]. Nonetheless, these "soluble" receptors possess conserved regions known to interact with the primary chemotaxis histidine kinase CheA, as well as determinants needed to form the distinctive trimer-of-dimers assembly state characteristic of transmembrane receptors. Indeed, soluble receptors organize into a hexagonal lattice bound to a baseplate of CheA and CheW at their membrane-distal tips, much like their transmembrane counterparts[1]. Nevertheless, what such receptors sense and how they convey signals to CheA is largely an open question.

A particular class of soluble receptor without a ligand-binding domain is found in the genomes of Thermotagates and is exemplified by the *T. maritima* (Tm) receptor TM14. TM14 is a thermostable, soluble receptor that has been biochemically and structurally characterized[2]. The structure of TM14 reveals the classic dimeric coiled-coil structure of MCPs that contains an adaptation region with putative sites for methylation by the methyl-transferase CheR and a signaling tip that interacts with CheA and the adaptor protein CheW. In reconstituted systems, TM14 deactivates CheA autophosphorylation activity[3]. Although TM14 is known to be a receptor that influences CheA activity, its role in chemotaxis remains to be understood. In general, evidence suggests that cytoplasmic receptors respond to the metabolic state of the cell, but it is

unknown what signals or ligands the receptors recognize[4]. Intriguingly, the tm14 gene clusters closely with another gene, tm13, which is atypical of previously studied chemotaxis proteins but belongs to a superclass of enzymes known to react with a wide range of substrates, including antibiotics and nitric oxide.

Bioinformatics analysis of bacterial genomes described below reveals that homologs of TM13 and TM14 are conserved as gene neighbors in 6 phyla of bacteria, including the pathogenic spirochete *T. denticola* (Td) (Fig. 1). Herein, we demonstrate that TM13 is a di-iron binding protein that functions in the chemotaxis of an example organism of *T. denticola*. These data, taken with additional results from biochemistry and protein crystallography, establish TM13 (hereafter named ODP1 for Oxygen Di-iron Protein) as a chemotactic oxygen and iron sensor that likely regulates the activity of soluble chemoreceptors.

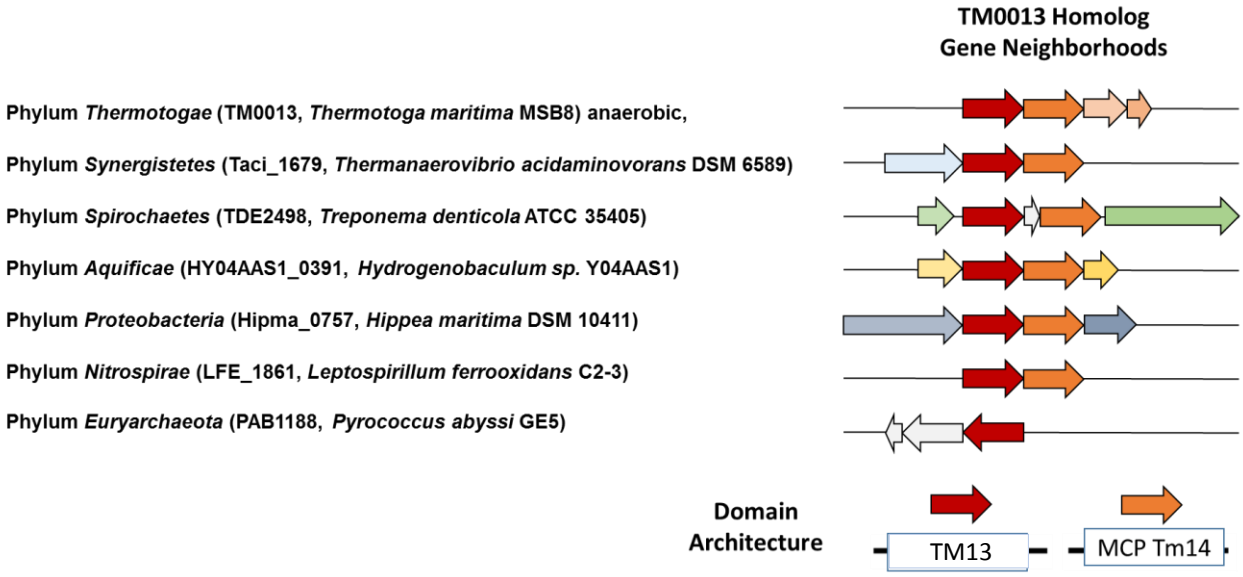


Figure 1: Bioinformatics analyses of bacterial genomes reveals that homologs of ODP1 are conserved as a gene neighbor to soluble receptors is six phyla of bacteria. Homologs of ODP1 are in red and homologs of the soluble receptor Tm14 are in orange. None of the other adjacent genes are conserved across the phyla.

3.2 Results

3.2.1 Characterized homologs of TdODP1 and TmODP1

TdODP1 and its closest related homologs are currently structurally and biochemically uncharacterized. However, some distant related homologs of TdODP1 have been studied and fall into two families, β-lactamases and flavo-diiron proteins (FDPs). β-lactamases are bacterial proteins responsible for resistance to β-lactam antibiotics[5]. These proteins are zinc-hydrolases that catalyze the hydrolysis of antibiotics to non-lethal substances. β-lactamases have ~20% sequence identity to TdODP1. FDPs are iron and flavin-binding proteins that either reduce oxygen (O_2) or nitric oxide (NO) to avoid oxidative and nitrosative stress, respectively[6]. These proteins are comprised of two domains: a metal-binding domain containing two iron atoms and a flavin-binding domain. Catalysis is accomplished by the oxidation of the iron atoms and flavin to reduce O_2 and NO. TdODP1 only contains the iron-binding domain, which has ~25% sequence identity to the FDP proteins. Sequence alignments of TdODP1 to each homolog family suggests that TdODP1 contains the necessary residues to bind both zinc and iron, as the metal-binding residues of the protein classes are similar.

3.2.2 Purification and metal analysis of TdODP1

TdODP1 can be expressed and purified in relatively high abundance in *E. coli* BL21 cells. The purified protein is purple in color and has a broad absorbance spectrum with two peaks at 500 nm and 590 nm (Fig. 2). Reduction of TdODP1 using dithionite diminishes the 500 nm/590 nm peaks and partial recovery of the spectra is achieved by re-exposure to air. Metal analysis by ICP-ES reveals that iron is bound to the protein monomer in a ~2:1 ratio with no other metals being found in significant abundance (Fig. 2A). These data are consistent with the fact that the closest related TdODP1 homologs are diiron proteins.

A

Ion	Protein:Ion ratio
Fe	1 : 1.8
Zn	1 : 0.06
Ni	1 : 0

B

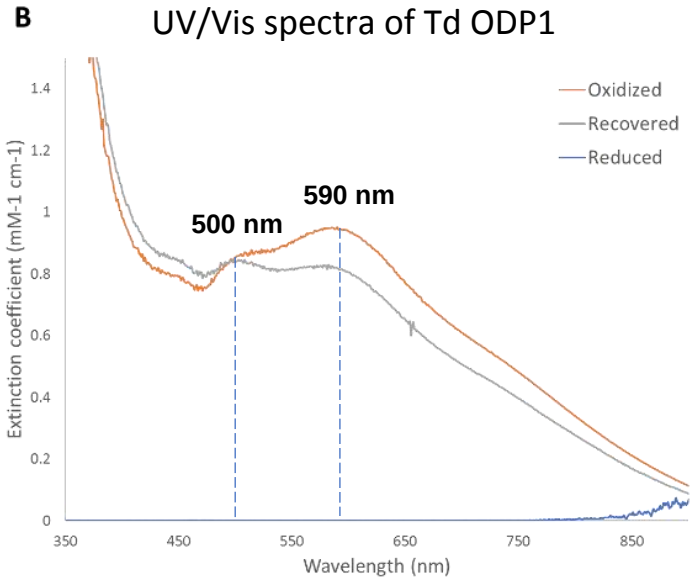


Figure 2: (A) Purified Td ODP1 is purple in color. Metal analysis by ICP-ES identifies that iron is present in the protein sample in a 2:1 ratio of protein to metal. (B) UV/Vis experiments show that freshly purified Td ODP1 has an absorbance maxima at 590 nm. Reduction with ditionite eliminates the purple color and other spectral features. Re-exposure to oxygen partially recovers the 590 nm absorption band and associated color.

3.2.3 Purification and metal analysis of Tm ODP1

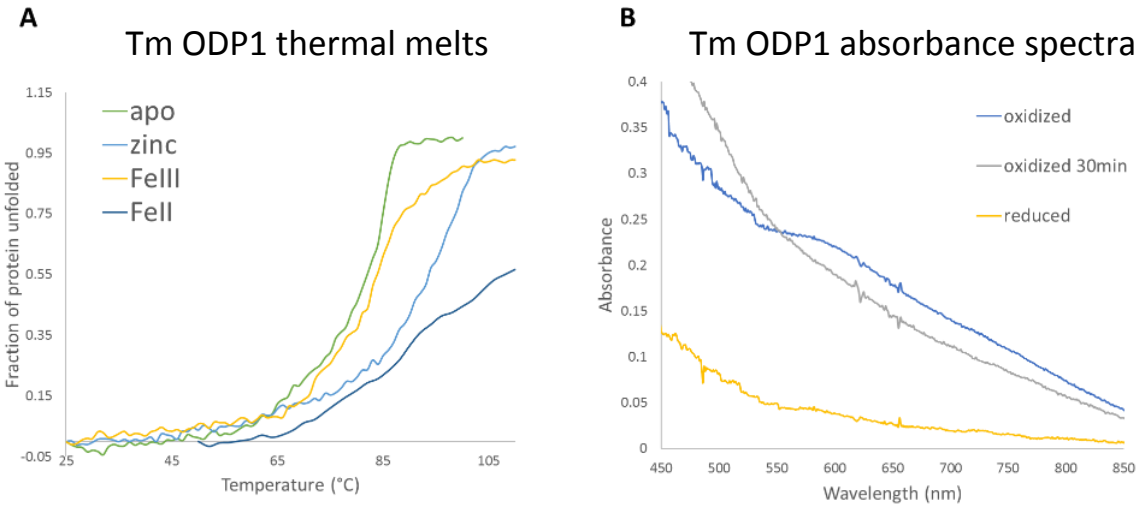


Figure 3: (A) Thermal melts were used to identify the native metal ligand of Tm ODP1 since its purified in its apo state. The apo Tm ODP1 has a Tm of 83°C. Reconstitution with zinc increases the Tm to 93°C. However, reconstitution with Fe(II) results in Tm of ~100°C. (B) Like Td ODP1, Tm ODP1 (Fe) has an absorbance maxima at 590 nm.

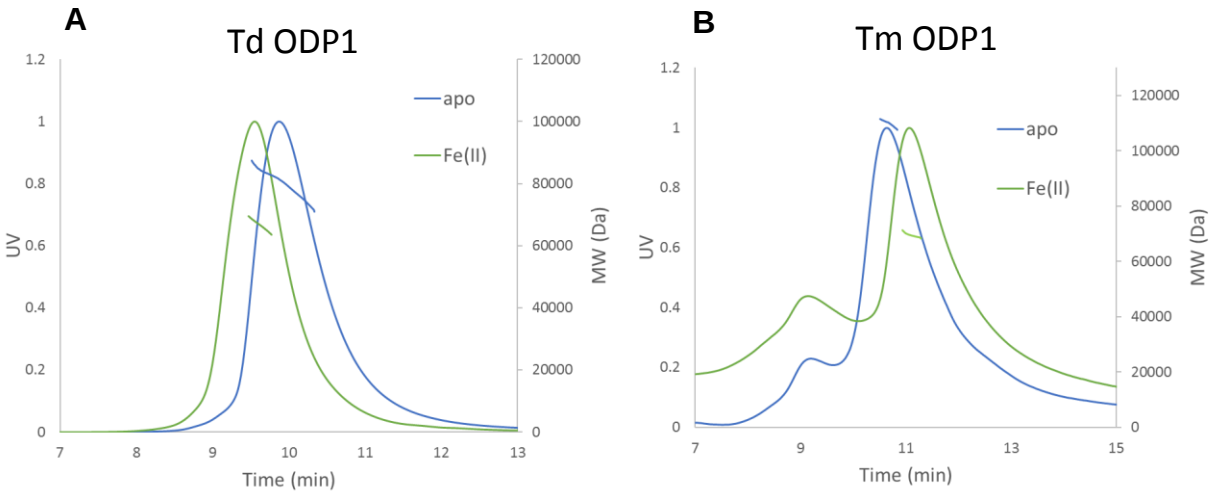


Figure 4: (A) MALS experiments with Td ODP1 demonstrate a molecular weight shift that occurs when iron is removed from the protein. (B) Likewise, the average molecular weight of Tm ODP1 increases when its present in its apo state.

TmODP1 can be expressed and purified in relatively high abundance E. coli BL21 cells. Unlike TdODP1, TmODP1 is colorless in solution after purification. Furthermore, metal analysis by ICP-ES reveals that TmODP1 is purified without reconstituted metal. This discrepancy between

the homologs is likely because TmODP1 originates from an extremophile (organism grows at 90 °C, 2 atm) and therefore does not possess native behavior under standard conditions.

Since TmODP1 is purified from cell lysate in the apo form, thermal melts were used to determine the native metal ligand. Based upon the activity of characterized homologs, iron and zinc were tested as candidate ligands. Although zinc and Fe(III) increase the melting temperature of the apo protein, TmODP1 is most stable when Fe(II) is reconstituted into the metal site (Fig. 3A). Furthermore, oxidation of the Fe(II)-reconstituted protein results in a color and absorbance spectra similar to that of the Td ODP1 protein (Fig. 3B).

3.2.4 MALS of TdODP1/TmODP1

Multi-angle light scattering (MALS) experiments show that TdODP1 forms a dimer in solution when the iron-peroxo adduct is present (60 kDa). However, the average molecular weight of the protein particles increases when the iron is removed (85 kDa) (Fig. 4A). These data suggests that apo TdODP1 undergoes a shift in the average oligomeric state, likely due to decreased stability of the protein upon ligand removal. Furthermore, thermal melts of TdODP1(Fe) and apo TdODP1 reveal a slight decrease in the melting temperature (2°C) of the apo protein.

MALS experiments with apo TmODP1 and metal-reconstituted TmODP1 demonstrate that, like the TdODP1 homolog, the average molecular weight of TmODP1 increases in the apo state. Specifically, apo TmODP1 forms a 120 kDa tetramer, while the metal-reconstituted proteins are predominantly a protein dimer (Fig. 4B).

3.2.5 Resonance Raman of Td ODP1

The absorbance spectra of the purple ODP1 suggest that the di-iron site contains an iron-peroxo adduct[7]. UV/Vis spectroscopy confirmed the characteristic 500 nm and 590 nm peaks were due to the binding of oxygen to Td ODP1; air-free oxygenated buffer was titrated into

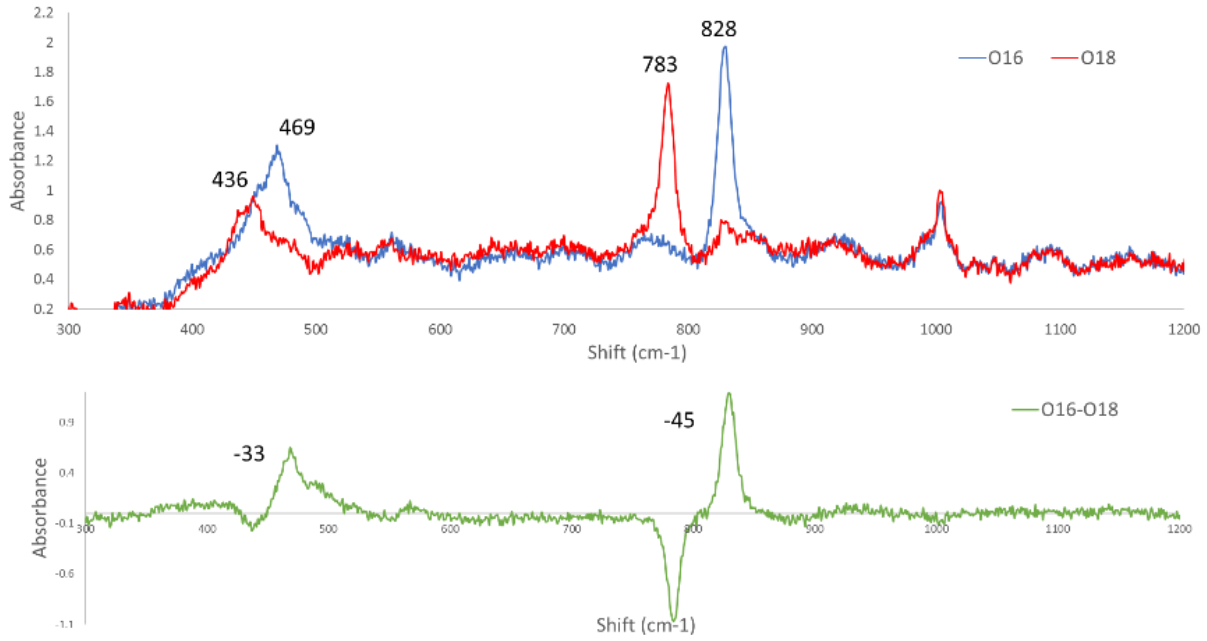


Figure 5: (A) Resonance Raman experiments with Td ODP1 demonstrate the presence of an iron peroxo species in the protein. The raman shifts were confirmed to be dependent on the presence of oxygen via isotope shift assays with O^{18}. (B) Difference spectra from the O^{16} and O^{18} experiments are indicative of a cis-μ-1,2 peroxo species.

reduced TdODP1 under anaerobic conditions. Resonance Raman (RR) experiments were conducted to determine if Td ODP1 does possess an iron-peroxo species (Fig. 5). Comparison of the RR spectra to previously characterized biomimetic molecules demonstrates the presence of a symmetric iron-peroxo adduct in aerobic Td ODP1[7]. Upon isotropic substitution of O_2 with O_2^{18}, both the band shifts of 469 cm^{-1} and 828 cm^{-1} bands shift to a lower energy as is expected for an Fe(III)-O_2^{2-} species. Reduction of Td ODP1 with dithionite eliminates the bands entirely.

3.2.6 Td ODP1 oxygen-binding is reversible

Iron-peroxo adducts are commonly utilized by proteins to catalyze oxygen-dependent redox reactions. In these cycles, the two Fe(II) ions bind oxygen and are oxidized to Fe(III) during catalysis. If the iron-peroxo species does not react with its substrate, it will spontaneously

oxidize to an Fe(III) species[8]. In both outcomes, the protein cannot form the active Fe(II) species once again until it has been reduced by a cofactor or coenzyme.

Unique to all other characterized non-heme iron-peroxo proteins, ODP1 spontaneously releases oxygen from the iron center under anaerobic conditions to return to the Fe(II) state that can again bind oxygen to form the iron-peroxo adduct. The ability to reversibly bind oxygen, and return to the Fe(II) active state, may allow ODP1 to act as an oxygen sensor that does not need a coenzyme or cofactor for multiple binding-and-release events. The ability of ODP1 to reversibly bind oxygen is demonstrated by incubating the protein solution in anaerobic conditions and then re-exposing the protein sample to oxygen (Fig. 6).

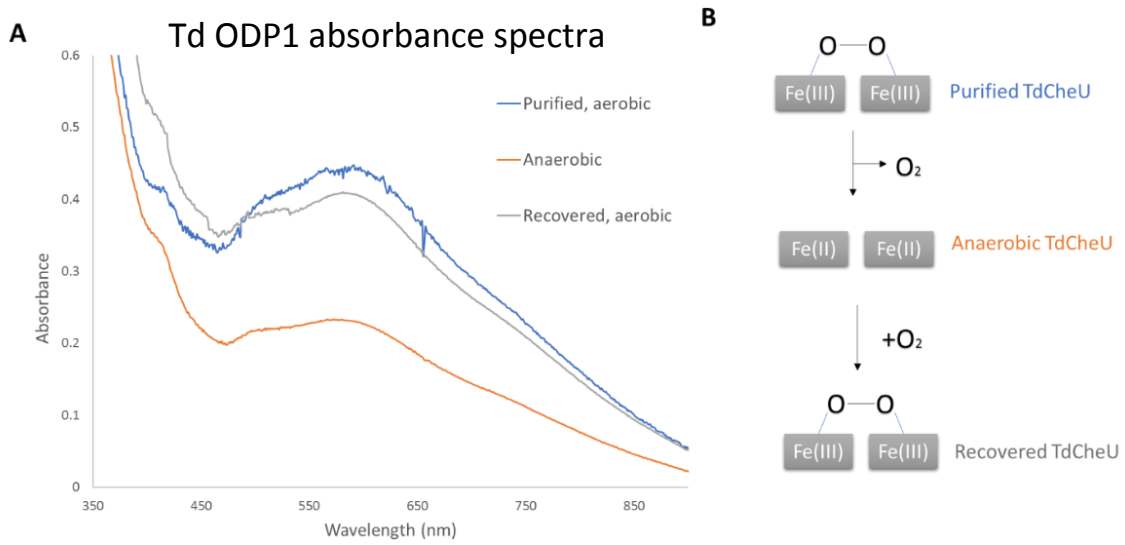


Figure 6: (A) UV/Vis experiments that monitor the absorbance at 590 nm demonstrate that Td ODP1 reversibly binds oxygen. When the protein is incubated in an anaerobic environment, the relative amount of iron-peroxo present decreases. However, when the protein is re-exposed to oxygen, the absorbance at 590 nm increases once again. (B) A schematic of oxygen release and binding by Td ODP1. The protein reverts back to an Fe(II) state after oxygen release, a unique feature of Td ODP1 and potential mechanism for oxygen sensing.

3.2.7 The iron-peroxo adduct is extremely stable in ODP1

A majority of iron-peroxo adducts found in proteins are extremely short-lived intermediates that are difficult to isolate due to their instability. Some rare exceptions include the human deoxyhypusine hydroxylase enzyme, a protein that regulates cell growth, which has an half-life of 30 hours at 20°C[9,10]. Similarly, under aerobic conditions the O_2 adduct of ODP1 is extremely stable with an apparent half-life of ~50 hours in Td and ~90 hours in Tm at 25°C (Fig. 7). This dramatic increase in stability compared to other family members is likely due to the hydrophobic nature of the oxygen-binding pocket, a hallmark of hyper-stable diiron-peroxo proteins, and the presence of a glutamine residue that is not present in the canonical diiron binding site in flavo-diiron proteins. Although the UV/Vis data suggests that the ODP1 iron-peroxo species is extremely long-lived in aerobic conditions, ODP1 can reversibly bind oxygen so the apparent half-life may reflect several O_2 binding-and-release events and is therefore not a true calculation for the half-life of a single iron-peroxo adduct. Instead, the calculation may represent the time is takes for the di-iron center to degrade due to iron release or protein unfolding.

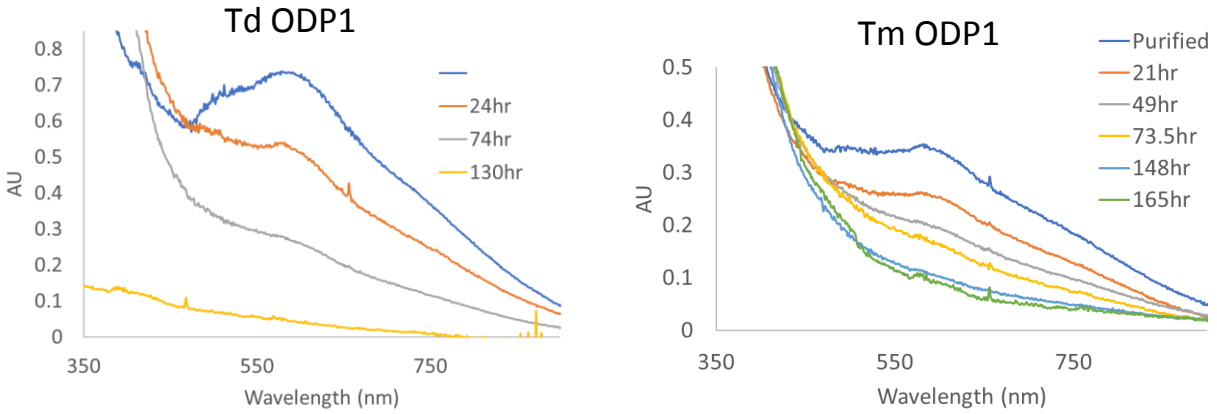


Figure 7: UV/Vis experiments that monitor the absorbance at 590 nm were used to calculate the apparent half-life of the iron-peroxo species of Td and Tm ODP1. (A) The half-life of Td ODP1 is 50 hours 25°C and (B) the half life of Tm ODP1 is 90 hours at 25°C.

3.2.8 Td ODP1 releases oxygen at a rapid rate

In the reversible oxygen-binding assays conducted above, the oxygen was slowly released from the protein sample in an anaerobic chamber over 10 hours. These results suggest that ODP1

binds the associated oxygen ligand with high affinity. However, it's possible that rapid dissociation from oxygen does occur, but that the released oxygen quickly re-associates with an iron center before it can diffuse out of solution, making long incubations under anaerobic conditions necessary for complete oxygen removal.

To determine if oxygen release does occur rapidly from ODP1, glucose oxidase-catalase system[11] was introduced to scavenge released oxygen and spectral changes associated with O_2 release were followed. Glucose oxidase scavenges free oxygen and glucose in solution to synthesize a glucose derivative (Fig. 8A). The structure of iron-reconstituted TdODP1 reveals that the oxygen ligand is buried in the active site, and therefore is not accessible to glucose oxidase while associated with TdODP1. However, once the oxygen is released from ODP1, glucose oxidase will consume the oxygen so that it cannot bind with another ODP1 subunit and the 590 nm absorbance peak of TdODP1 will diminish.

The results of this experiment demonstrate that the 590 nm peak of TdODP1 is completely diminished after 90 seconds of exposure to the glucose oxidase-catalase system, which means that oxygen is rapidly released from TdODP1 (Fig. 8B). Therefore, the apparent stability of the iron-peroxo species in ODP1 does not reflect the stability of a single oxygen adduct, but reflects the ability of the protein to continuously bind and release oxygen over time with high affinity.

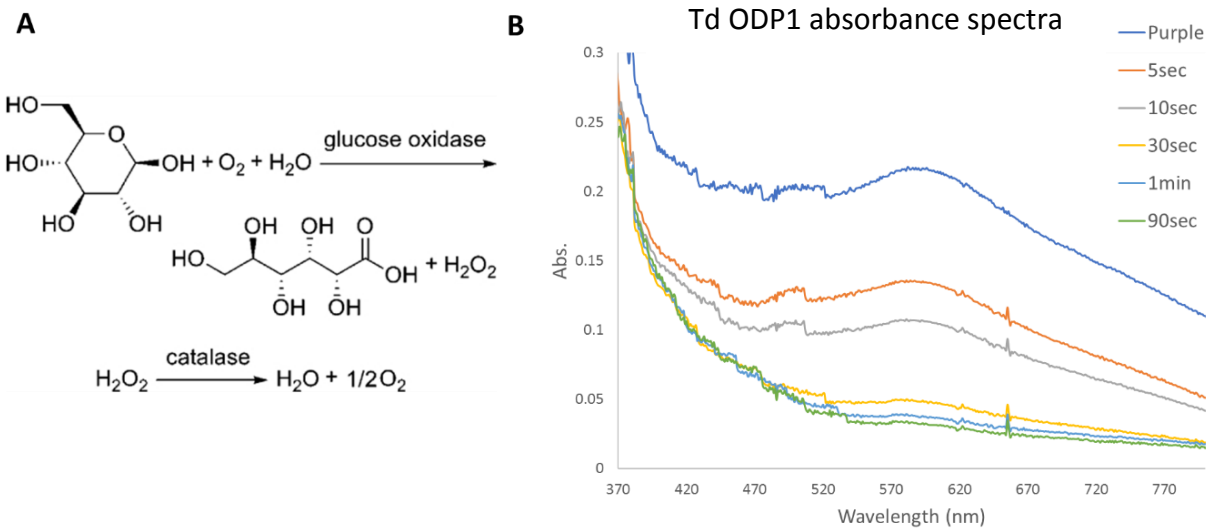


Figure 8: (A) The glucose oxidase/catalase system is used to remove accessible oxygen from solution. (B) Addition of the glucose oxidase/catalase system to a sample of Td ODP1 demonstrates that the oxygen is rapidly released from the protein. After 90 seconds, all the detectable iron-peroxo (590 nm) was diminished.

3.2.9 In vivo chemotaxis assays

The biochemical experiments presented demonstrate that ODP1 has two ligands, iron and oxygen. To test if ODP1 is responsible for sensing these ligands as part of a chemotactic response, chemotaxis capillary assays were conducted in *T. denticola*. A ODP1 knock-out strain (Δ2498) was generated and compared to the behavior of a wild-type (WT) cell line.

These strains were then used in attractant-based capillary assays with two candidate iron sources, iron and hemin. Previous experiments have identified hemin as source of iron for *T. denticola.* When hemin is used as a chemoattract, a significant decrease in chemotaxis is seen

for the Δ2498 cell line. Additionally, the use of Fe(II) as an attractant results in complete loss of chemoattraction in Δ2498. Together, these data confirm that TdODP1 is an iron sensor, and that free Fe(II) and hemin are chemo-attractants for this organism (Fig 9). As a positive control, the ability of the Δ2498 cell line was tested for chemo-attraction to glucose[11] to ensure that deletion of ODP1 does not impair chemotaxis in general; compared to the WT cell line, no significant difference was found in the Δ2498 cell line.

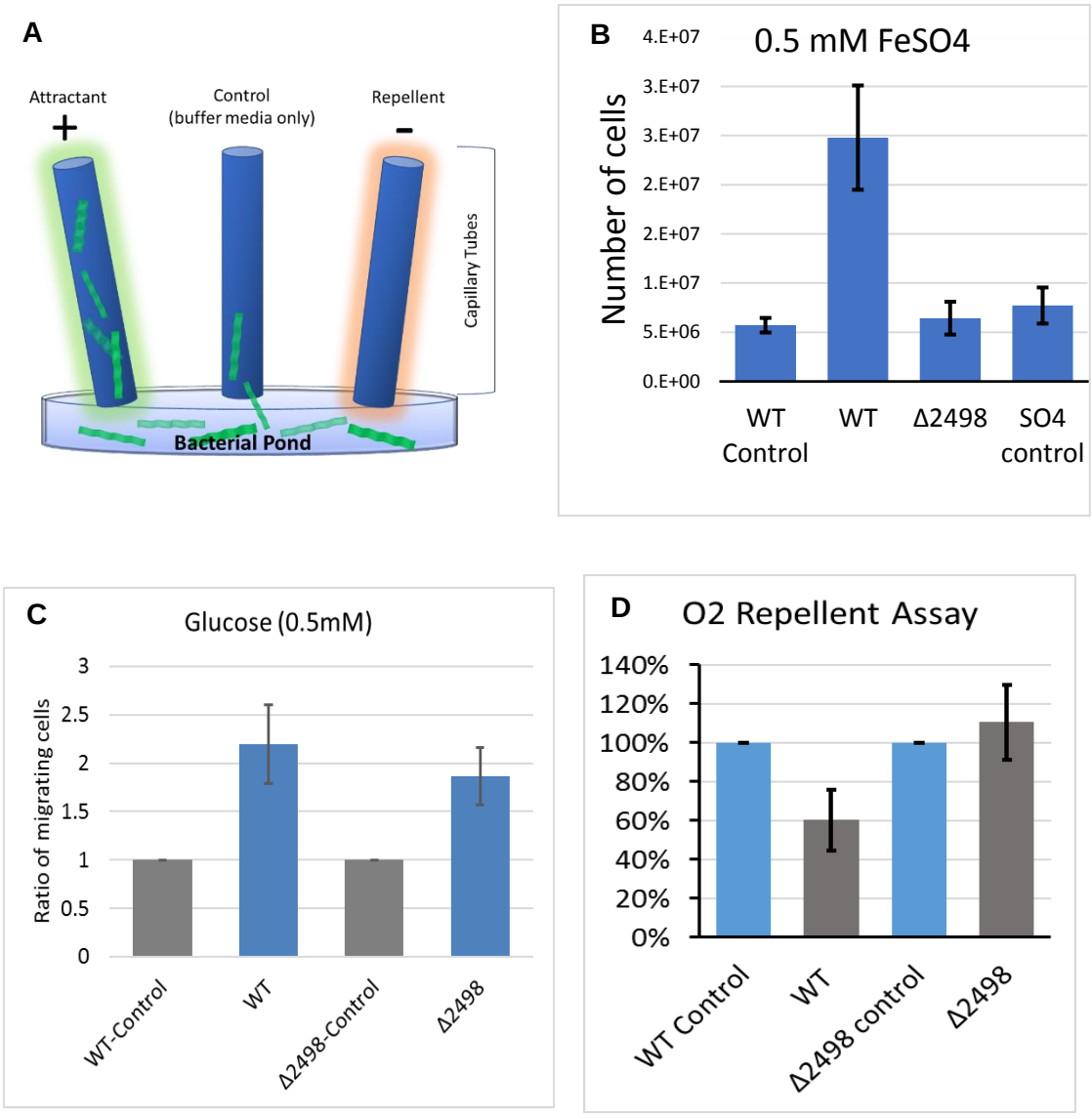


Figure 9: In vivo capillary assays with wild-type and Td ODP1 knock out (Δ2498) cell lines were used to determine if Td ODP1 is responsible for sensing Fe(II) and oxygen *in vivo*. (A) An illustration for the *in vivo* capillary assays. *T. denticola* cells swim toward attractants or away from repellents (B) Capillary assays using Fe(II) as an attractant verify that Td ODP1 is an iron sensor. (C) Glucose control experiments validate that the Δ2498 cell line can still chemotax toward glucose to the same extent as the wild-type cell line. (D) Repellent assays with oxygenated buffer verify that Td ODP1 is an oxygen sensor. Data and figures were contributed by Dr. Chunhao Li.

Repellent-based capillary assays using oxygen as a candidate repellent were used to determine

if ODP1 also functions as an oxygen sensor. Since the 6 phyla possessing a receptor-coupled ODP1 homolog are anaerobes, they likely possess a system for recognizing and avoiding oxygen. Indeed, when ODP1 is not present in the cells, there are significantly more Td cells in capillary tubes, indicating that the cells can no longer sense and swim away from aerobic environments (Fig. 9).

3.2.10 Crystal structures of metal-reconstituted TdODP1/TmODP1

The 2.07 Å resolution crystal structure of TdODP1 reveals a protein dimer with a diiron active site, as confirmed by ICP-ES. Like its characterized homologs, the active site contains 7 ligand residues, with one aspartate residue bridging both metal ions (Fig 10)[6,12]. Furthermore, the electron density around the active site reveals the presence of the cis-μ-peroxo species that was identified by resonance raman experiments.

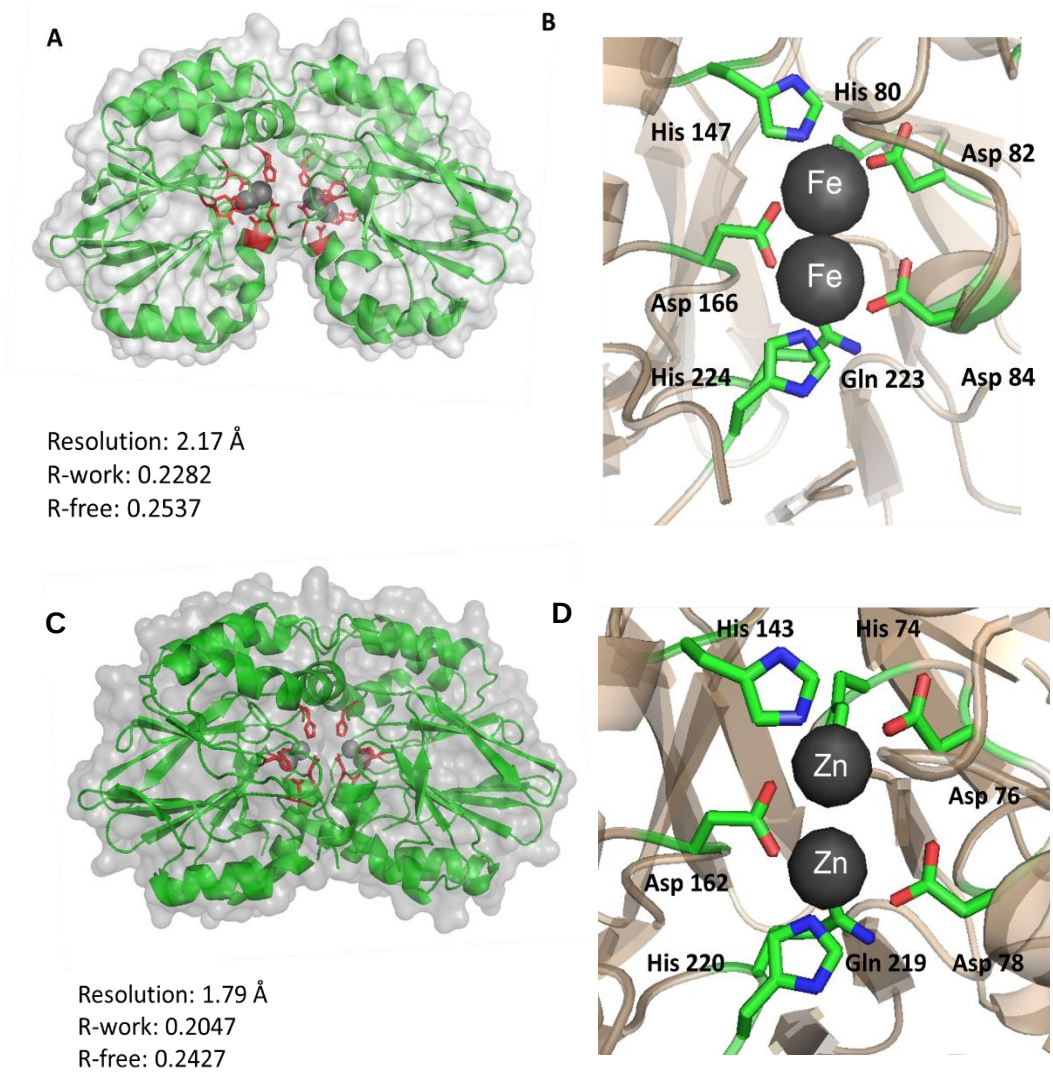


Figure 10: The crystal structures of metal-reconstituted Td and Tm ODP1 reveal the presence of the protein dimer with a unique di-metal binding site. (A) A ribbon model of Td ODP1 (Fe) with the transparent surface showing. Metal-binding residues are highlighted in red and the two irons are shown as grey spheres. (B) The ligand binding site of Td ODP1 possesses a unique glutamine residue (Gln 223) that is normally conserved as histidine in ODP1 homologs. (C) A ribbon diagram of Tm ODP1 (Zn) with the transparent surface showing. Metal-binding residues are highlighted in red and the two irons are shown as grey spheres. (D) Like Td ODP1, the ligand binding site of Td ODP1 possesses a unique glutamine residue (Gln 219) in the metal-binding site.

Since TmODP1 does not stay bound to iron under most buffer conditions, TmODP1 was

crystallized after reconstitution with zinc. The 1.8 Å resolution zinc-reconstituted structure reveals the protein dimer calculated by MALS. In order to determine the structure of TmODP1 reconstituted with Fe(II), TmODP1(Zn) crystals were put into an anaerobic chamber and a solution containing 10mM Fe(II)SO_4 was added to the drops. Although the structures of TmODP1(Zn) and TmODP1(FeII) are very similar, the resolution of the Fe(II) crystals are greatly reduced (3.24 Å resolution). It's likely that the Fe(II) has oxidized to Fe(III) during crystal freezing and data collection; TmODP1 does not stay bound to Fe(III) and returns to its tetrameric apo state, which destabilizes crystal contacts and decreases resolution.

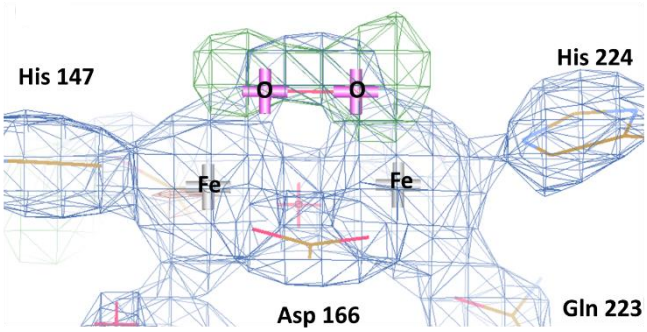


Figure 11: The electron density maps of the Td ODP1 crystal structure shows that the active site of TdODP1 contains the cis-μ-peroxo species identified by Resonance Raman experiments. Blue density: 2Fo-Fo at σ 2.3. Green density: Fo-Fc at σ 3.3.

3.2.11 Crystal structures of apo TdODP1/TmODP1

The 2.2 Å resolution crystal structure of apo TdODP1 is similar to the iron-reconstituted structure with the exception of 6-residue loop that contains metal-binding residues. In the apo structure, these residues have moved further away from the iron-binding site, resulting in greater solvent accessibility of the binding pocket. The resulting increase in solvent accessibility may be a mechanism that allows for iron sensing (Fig. 11).

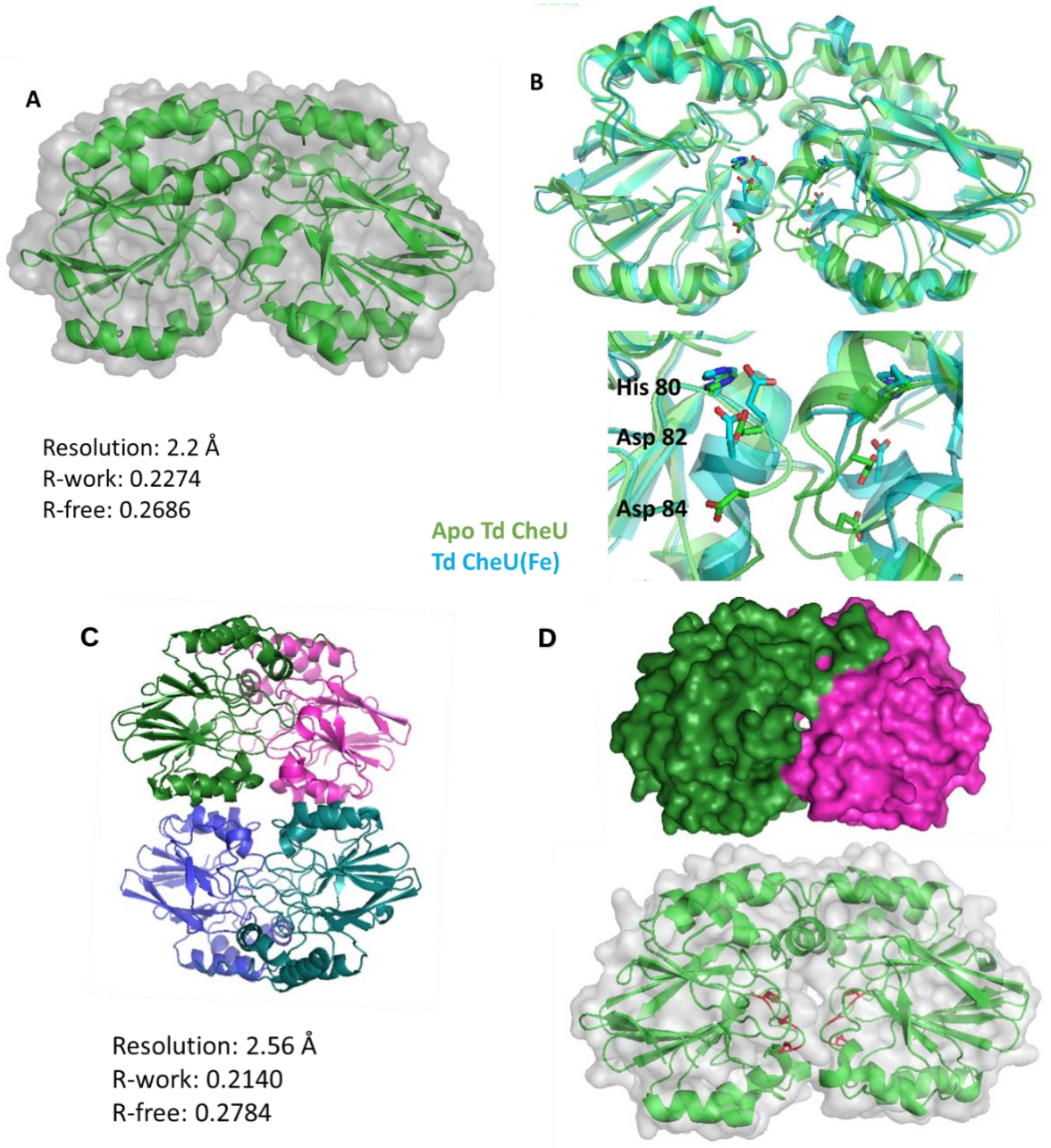


Figure 12: (A) The crystal structure of apo Td ODP1 shown as a ribbon model with the transparent surface showing. (B) An overlay of Td ODP1 (Fe) (cyan) and apo Td ODP1 (green) reveals conformational changes at the metal-binding site. (C) The crystal structure of apo Tm ODP1 as a ribbon model reveals the protein tetramer. (D) The surface of apo Tm ODP1 reveals a 3.5 Å ion channel that bypasses the metal-binding residues highlighted in red.

The 2.56 Å resolution crystal structure of apo TmODP1 reveals the protein tetramer calculated by MALS. Furthermore, examining the solvent-exposed surface of the tetramer reveals a putative solvent channel that bypasses the metal-binding residues (Fig. 12). This conformational change, like those seen in the Td homolog, may be a mechanism for binding iron present in the cytosol.

3.2.12 Testing ODP1 activity *in vitro*

Collectively, these data demonstrate that ODP1 is an oxygen and iron sensor for the chemotaxis system. While bioinformatics analyses suggest that ODP1 interacts with soluble receptors to influence the chemotaxis machinery, *in vitro* reconstitution of the chemotaxis system has not demonstrated a direct interaction among ODP1 and other known chemotaxis proteins (soluble receptors, CheA, CheW, CheR, CheY). Radioisotope assays that monitor CheA and CheY phosphorylation, protein:protein pull-down assays, cross-linking assays and UV/Vis experiments have not been able to demonstrate that ODP1 influences CheA activity or directly interacts with any of the listed proteins. This discrepancy may due to the fact that we are not able to properly reconstitute the chemotaxis system *in vitro.* The complex oligomeric state of receptors and supramolecular assembly of chemotaxis proteins, which is necessary for proper *in vivo* function, make the system difficult to recapitulate *in vitro*. Further research is being conducted to properly reconstituted this system and determine the mechanism for chemotactic sensing by ODP1.

3.3 Discussion

The sensing abilities and mechanisms for soluble chemoreceptors have remained elusive despite advancements in the characterization of membrane-bound arrays. Difficulties in investigating the soluble system arise from the fact that the chemoreceptors have no obvious sensor domain and *in vitro* reconstitution of intact functional arrays are challenging to produce.

However, recent data supports the hypothesis that the soluble systems are responsible for monitoring the current metabolic state and intracellular environment of the cell.

In these experiments, we demonstrate that an intracellular protein that genetically clusters with a soluble chemoreceptor is responsible for oxygen and iron sensing. Interestingly, previous *in vivo* experiments with wild-type *T. denticola* (Td) cells had established that Td recognizes hemin and oxygen as an attractant and repellent, respectively[13,14]. However, these experiments did not pursue identification of the chemotaxis protein(s) involved in recognizing these ligands. Here, we demonstrate that deletion of ODP1 completely eliminates chemotactic recognition of oxygen and iron. This protein, ODP1, is not excreted from the cell and therefore presumably monitors the intracellular levels of these ligands. The protein structures and biochemical analyses shown here support a model where the presence of iron induces conformational changes in TmODP1/TdODP1, which ultimately leads to fluctuations in protein stability and oligomeric state. However, how the protein ODP1 interacts with the soluble receptor to influence chemotaxis remains unclear. To answer this question, further research must be done to properly reconstitute the soluble arrays *in vitro* so that the isolated system may be biochemically characterized.

3.4 Methods

3.4.1 Cloning and purification of TmODP1/TdODP1

ODP1 from T. maritima and T. denticola was cloned into pet28 and transformed into *E. coli* BL21 (DE3) cells. After plating the transformation on agar containing kanamycin, a single colony was chosen for cell culture growth. 8 L of cells were grown at 37°C until an O.D. at 600 nm reached 0.6. 1 mM IPTG was added to each flask at the cells were grown for 16 hours at room temperature. After the cells were pelleted via centrifugation, they were resuspended in 50 ml of lysis buffer (50 mM MOPS pH 7.5, 150 mM NaCl, 5 mM Imidazole) and sonicated for 6 minutes.

The lysed cells were then centrifuged at high speed to remove the insoluble fraction from the lysate. The lysate was then run over a Ni-NTA affinity resin, washed with 20 mM Imidazole buffer to remove non-specific binding proteins, and finally eluted using buffer with 200 mM imidazole. The eluted protein was further purified by size-exclusion chromatography (Sephadex s200). The fractions containing the protein of interested were pooled and concentrated to ~ 500 µM.

3.4.2 ICP-ES

ICP-ES experiments were conducted on purified Td and Tm ODP1 to determine the concentration of various metals in the protein sample. The samples were prepared in 5 ml of buffer (50 mM Tris pH 7.5, 150 mM NaCl) with a total protein concentration of 1 µM.

3.4.3 Metal reconstitution

When Tm ODP1 is purified from *E. coli* cells, it is present in its apo state. Iron reconstitution into Tm ODP1 is accomplished by placing 3 ml of 100 µM purified Tm ODP1 (50mM MOPS pH 7.5, 150 mM NaCl) in an anaerobic chamber for 3 hours to remove the oxygen from solution, and then adding four-times molar excess $FeSO_4$ to the protein solution. After incubation for 1 hour, excess iron is removed from the solution via buffer exchange by either SEC or several rounds of dilution followed by concentration. Zinc was reconstituted into Tm ODP1 via dialysis against 2L of buffer containing 10µM $ZnSO_4$, 10 mM Sodium Cacodylate, 150mM NaCl followed by buffer exchange to remove excess zinc.

3.4.4 Thermal melts

Thermal melts via circular dichroism were conducted to determine the native ligands of Tm ODP1. For each sample, the unfolding of 20µM of purified Tm ODP1 was monitored by measuring the absorbance at 220 nm, which corresponds to the absorbance of alpha helices, from 25°C to 110°C.

3.4.5 Multi-angle light scattering (MALS)

Size-exclusion chromatography coupled to multi-angle light scattering experiments were used to determine the oligomeric state of apo and iron-reconstituted ODP1. For each sample, 5mg/ml of protein was injected onto a Phenomenex reverse-phase column pre-equilibrated with 50mM MOPS pH 7.5, 150 mM NaCl at room temperature. Bovine serum albumin (sigma) was used as a protein standard. Wyatt technologies ASTRA 6 program was used for data analysis and molecular weight calculation.

3.4.6 Resonance Raman

Resonance Raman experiments were used to confirm the presence and geometry of an iron-peroxo adduct in Td ODP1. 300 µl of aerobic (O^{16}) 800 µM Td ODP1 was placed into a NMR tube and the absorbance at 638 nm with a 16 mW Neon/Helium diode laser (Thor labs, model HNL210L) was measured and averaged for 2 hours. O^{18} experiments were carried out by first reducing the O^{16}-incorporated protein with dithionite and then adding O^{18}-saturated buffer to the protein; the RR data was collected using a sealed Jung NMR tube to keep O^{16} from entering the sample. Experiments with reduced Td ODP1 were accomplished by reducing the protein with dithionite and carrying out data collection in a sealed Jung NMR tube.

3.4.7 In vivo experiments

Capillary assays were used to determine potential chemo-attractants and repellents for T. denticola. In these experiments, bacterial cells were harvested and re-suspended in microcentrifuge tubes, and capillary tubes containing candidate ligands were applied to the bacteria suspensions. Non-gradient controls were included where bacteria suspension contained the same concentration of candidate ligands. Cells from each capillary tube were then automatically counted and compared to the non-gradient controls to determine if the ligand influences chemotaxis.

3.4.8 Crystallography

Td ODP1 (Fe): TdODP1 (Fe) concentrated to ~15 mg/ml was crystallized via hanging drop vapor diffusion in a solution containing a 1:1 ratio of protein sample to 20% PEG 3500. Thin sheets of crystals are formed with 4-7 hours at room temperature. Crystals were manually picked up in loops, flash frozen, and shipped in liquid nitrogen to a synchrotron beamline (APS, line 24-ID-E). Diffraction data was first scaled and integrated using HKL2000, and phasing was done via molecular replacement through Phaser MR with the Tm ODP1 apo structure.

Table 3.1: Data collection and refinement statistics for TdODP1 (Fe)

Wavelength (Å)	0.97920
Synchrotron	APS
Beam-line	24-ID-E
Space group	$P2_12_12_1$
a, b, c	44.67, 85.45, 118.80
α, β, γ	90, 90, 90
Resolution (Å)	48.9 - 2.07
R_{merge}	0.151(1.51)
$R_{p.i.m.}$	0.062(0.725)
R_{meas}	0.164(1.687)
I/σ(I)	11.6 (1.0)
Completeness (%)	98.6 (84.7)
Multiplicity	7.0 (4.9)
Mosaicity	0.251
Total reflections	28082
Refinement	
Resolution (Å)	2.07
Reflections used for R_{free}	26582
R_{free}	0.2537
R_{work}	0.2282
Clash Score	18.75
R.M.S Deviations	
Bond lengths (Å)	0.027
Bond angles (°)	2.1
Ramachandran outliers (%)	0.97
Rotamer outliers (%)	5.45

Tm ODP1 (Zn): Tm ODP1 was reconstituted with zinc via dialysis in a solution of 10 mM Sodium cacodylate pH 7.0 with 5 µM zinc sulfate. The protein was concentrated to ~10mg/ml and crystallized via hanging drop vapor diffusion in a solution containing a 1:1 ratio of protein sample to 30% PEG 1500. Thin sheets of crystals were formed in 2-3 days. Crystals were manually picked up in loops, flash frozen and shipped in liquid nitrogen to a synchrotron beamline (CHESS, line F1). Diffraction data was first scaled and integrated using HKL2000, and phasing was done via molecular replacement through Phaser MR with the Tm ODP1 apo structure.

Table 3.2: Data collection and refinement statistics for TmODP1 (Zn)

Wavelength (Å)	0.9790
Synchrotron	CHESS
Beam-line	F1
Space group	P1
a, b, c	46.41, 65.12, 159.54
α, β, γ	89.98, 89.95, 110.88
Resolution (Å)	50 – 1.80
R_{merge}	0.078(1.174)
$R_{p.i.m.}$	0.083(1.112)
R_{meas}	0.158(2.064)
I/σ(I)	4.6 (3.3)
Completeness (%)	98.9(95.47)
Multiplicity	3.4(3.2)
Mosaicity	0.16
Total reflections	154882
	Refinement
Resolution (Å)	1.8
Reflections used for R_{free}	112634
R_{free}	0.2427
R_{work}	0.2047
Clash Score	21.64
	R.M.S Deviations
Bond lengths (Å)	0.011
Bond angles (°)	1.35
Ramachandran outliers (%)	0.49
Rotamer outliers (%)	0.78

Td ODP1 apo: Td ODP1 in the apo state was concentrated to ~14 mg/ml was crystallized via hanging drop vapor diffusion in a solution containing a 1:1 ratio of protein sample to 0.1 M Imidazole pH 7.0, 20% PEG 4000. Rectangular crystals formed within 1-3 days. Crystals were manually picked up in loops, flash frozen and shipped in liquid nitrogen to a synchrotron beamline (APS, line 24-ID-C). Diffraction data was first scaled and integrated using HKL2000, and phasing was done via molecular replacement through Phaser MR with the Tm ODP1 apo structure.

Table 3.3: Data collection and refinement statistics for TdODP1 apo

Wavelength (Å)	0.979200
Synchrotron	APS
Beam-line	24-ID-C
Space group	$P2_12_12_1$
a, b, c	45.22, 94.8, 117.09
α, β, γ	90, 90, 90
Resolution (Å)	49 - 2.02
R_{merge}	0.092(0.541)
$R_{p.i.m.}$	0.049(0.272)
R_{meas}	0.124(0.704)
I/σ(I)	4.6 (3.3)
Completeness (%)	98.9(95.47)
Multiplicity	6.4(6.6)
Mosaicity	0.283
Total reflections	63244
	Refinement
Resolution (Å)	2.02
Reflections used for R_{free}	33559
R_{free}	0.2686
R_{work}	0.2274
Clash Score	14.59
	R.M.S Deviations
Bond lengths (Å)	0.009
Bond angles (°)	1.3
Ramachandran outliers (%)	3.22
Rotamer outliers (%)	0.5

Tm ODP1 apo: Tm ODP1 in the apo state was concentrated to ~7 mg/ml was crystallized via hanging drop vapor diffusion in solution containing a 1:1 ratio of protein sample to 0.1 M Hepes pH 7.5, 7% Isopropanol, 23% PEG 4000. Rectangular crystals formed within 1-2 days and diffraction data was collected at a synchrotron beamline (CHESS). The diffraction data was first scaled and integrated using HKL2000. Phasing was accomplished via SAD phasing techniques with seleno-methionine substituted Tm ODP1.

Table 3.4: Data collection and refinement statistics for TmODP1 apo

Wavelength (Å)	1.0
Synchrotron	CHESS
Beam-line	F1
Space group	$P2_12_12_1$
a, b, c	84.65, 110.72, 110.96
α, β, γ	90, 90, 90
Resolution (Å)	50 – 2.56
R_{merge}	0.167(0.816)
$R_{p.i.m.}$	0.056(0.338)
R_{meas}	0.200(0.877)
I/σ(I)	4.6 (3.3)
Completeness (%)	97.8(95.47)
Multiplicity	11.6(6.1)
Mosaicity	0.58
Total reflections	33170
	Refinement
Resolution (Å)	2.56
Reflections used for R_{free}	33482
R_{free}	0.2140
R_{work}	0.2784
Clash Score	16.76
	R.M.S Deviations
Bond lengths (Å)	0.011
Bond angles (°)	1.40
Ramachandran outliers (%)	3.02
Rotamer outliers (%)	1.32

References

1. Briegel, A. *et al.* Structure of bacterial cytoplasmic chemoreceptor arrays and implications for chemotactic signaling. 1–16 (2014). doi:10.7554/eLife.02151

2. Pollard, A., Bilwes, A. & Crane, B. NIH Public Access. **40,** 1301–1315 (2012).

3. Greenswag, A., Muok, A., Li, X. & Crane, B. HHS Public Access. **427,** 87–92 (2016).

4. Wadhams, G. H. *et al.* TlpC, a novel chemotaxis protein in Rhodobacter sphaeroides, localizes to a discrete region in the cytoplasm. *Mol. Microbiol.* **46,** 1211–1221 (2002).

5. Palzkill, T. Metallo-β-lactamase structure and function. *Ann. N. Y. Acad. Sci.* **1277,** 91–104 (2013).

6. Kurtz, Jr., D. M. Flavo-diiron enzymes: nitric oxide or dioxygen reductases? *Dalt. Trans.* 4115 (2007). doi:10.1039/b710047g

7. Fiedler, A. T. *et al.* Spectroscopic and computational studies of (??-Oxo)(??-1, 2-peroxo)diiron(III) complexes of relevance to nonheme diiron oxygenase intermediates. *J. Phys. Chem. A* **112,** 13037–13044 (2008).

8. Krebs, C., Bollinger, J. M. & Booker, S. J. Cyanobacterial alkane biosynthesis further expands the catalytic repertoire of the ferritin-like ‘di-iron-carboxylate’ proteins. *Curr. Opin. Chem. Biol.* **15,** 291–303 (2011).

9. Makris, T. M. *et al.* An unusual peroxo intermediate of the arylamine oxygenase of the chloramphenicol biosynthetic pathway. *J. Am. Chem. Soc.* **137,** 1608–1617 (2015).

10. Han, Z. *et al.* Crystal structure of the peroxo-diiron(III) intermediate of deoxyhypusine hydroxylase, an oxygenase involved in hypusination. *Structure* **23,** 882–892 (2015).

11. Jd, R., Lux, R., Shi, W., Nw, C. & Effect, D. A. Effect of glucose on Treponema denticola

cell behavior. **35405,** 234–238 (2008).

12. Carfi, a *et al.* The 3-D structure of a zinc metallo-beta-lactamase from Bacillus cereus reveals a new type of protein fold. *EMBO J.* **14,** 4914–21 (1995).

13. Xu, X., Holt, S. C., Kolodrubetz, D. & Mmun, I. N. I. Cloning and Expression of Two Novel Hemin Binding Protein Genes from Treponema denticola. **69,** 4465–4472 (2001).

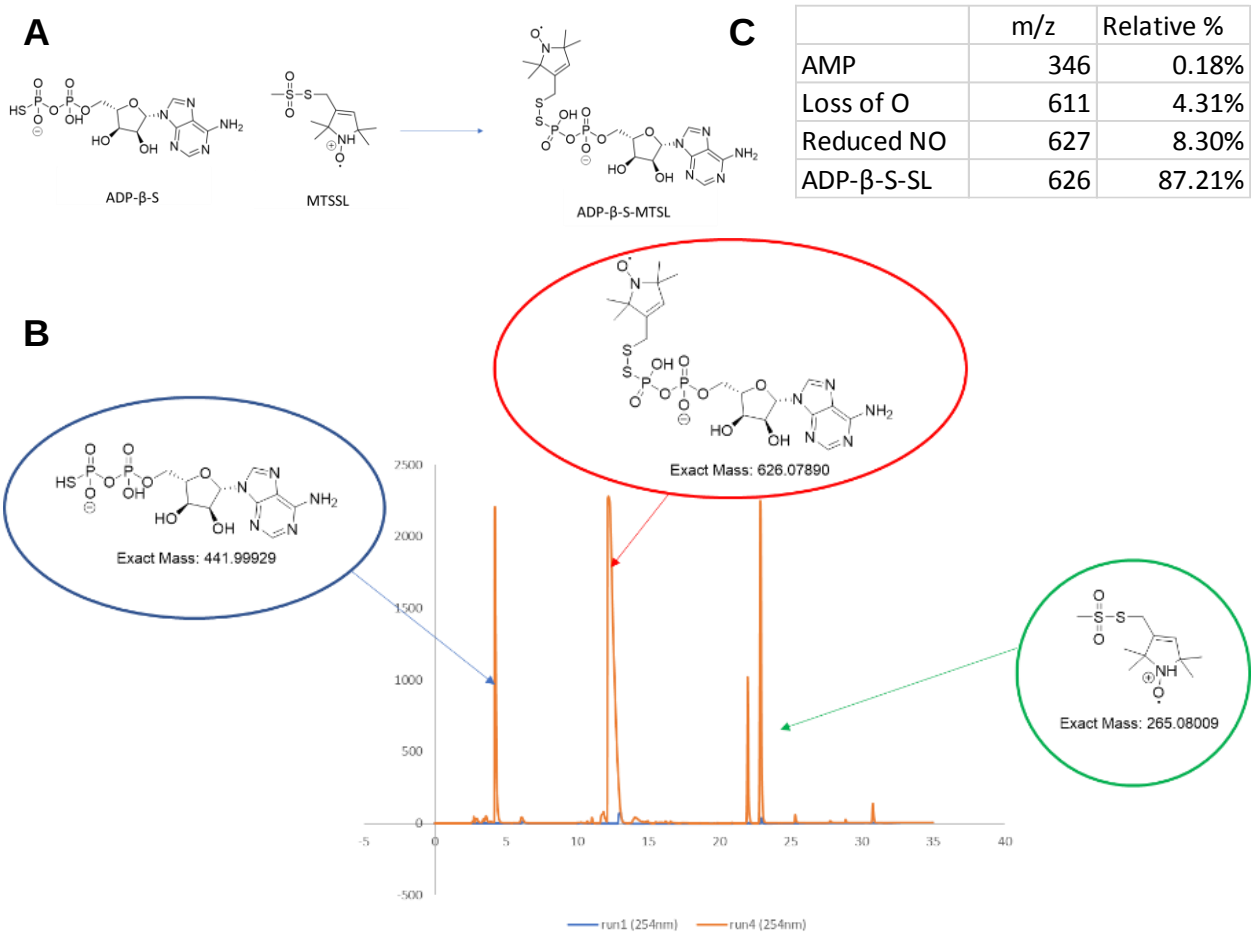
14. Lai, Y. & Chu, L. Novel Mechanism for Conditional Aerobic Growth of the Anaerobic Bacterium Treponema denticola . **74,** 73–79 (2008).

Chapter 4

A spin-labeling technique for ESR spectroscopy of ATP-binding proteins

4.1 Introduction

Site-directed spin labeling in concert with electron-spin resonance (ESR) spectroscopy is an invaluable tool for studying the nature of proteins in solution. This method has been especially useful for examining proteins that are flexible or highly dynamic; these properties make structural investigation otherwise difficult with techniques such as x-ray crystallography and small-angle x-ray scattering[1]. However, in the majority of cases, spin-labeling is achieved by chemically modifying engineered Cys residues with the small molecule MTSSL, which are ideally introduced at sites not essential for protein function. Such non-perturbative labeling may not always be possible when targeting sites near ligand-binding or protein:protein interaction regions. Furthermore, the protein may have many native Cys residues or otherwise be intolerable to residue substitution, genetic manipulation or recombinant expression. To circumvent some of these issues, alternative labeling approaches that do not rely on cysteine labeling have been developed. The use of genetically encoded synthetic amino acids has provided a means for spin-labeling proteins without the need for protein chemical modification or mutation of native cysteine residues[2,3]. However, these residue substitutions cannot be used in ligand-binding regions.



C

	m/z	Relative %
AMP	346	0.18%
Loss of O	611	4.31%
Reduced NO	627	8.30%
ADP-β-S-SL	626	87.21%

Figure 1: (A) The reaction scheme for the synthesis of ADP-β-S-SL. (B) The LC-MS trace and molecular weight calculations for the synthesis mixture. The peak corresponding to ADP-β-S-SL was subsequently isolated via prep-grade HPLC. (C) LC-MS analysis of the isolated ADP-β-S-SL after storage at -80°C and incubation at 25°C for 24 hours.

Previous methods have utilized the MTSSL spin-labeling method to label oligonucleotides after chemical modification of phosphates to a reactive phosphorothioate[4]. Synthesis of the spin-labeled oligonucleotide occurs through a similar mechanism as cysteine labeling and has been

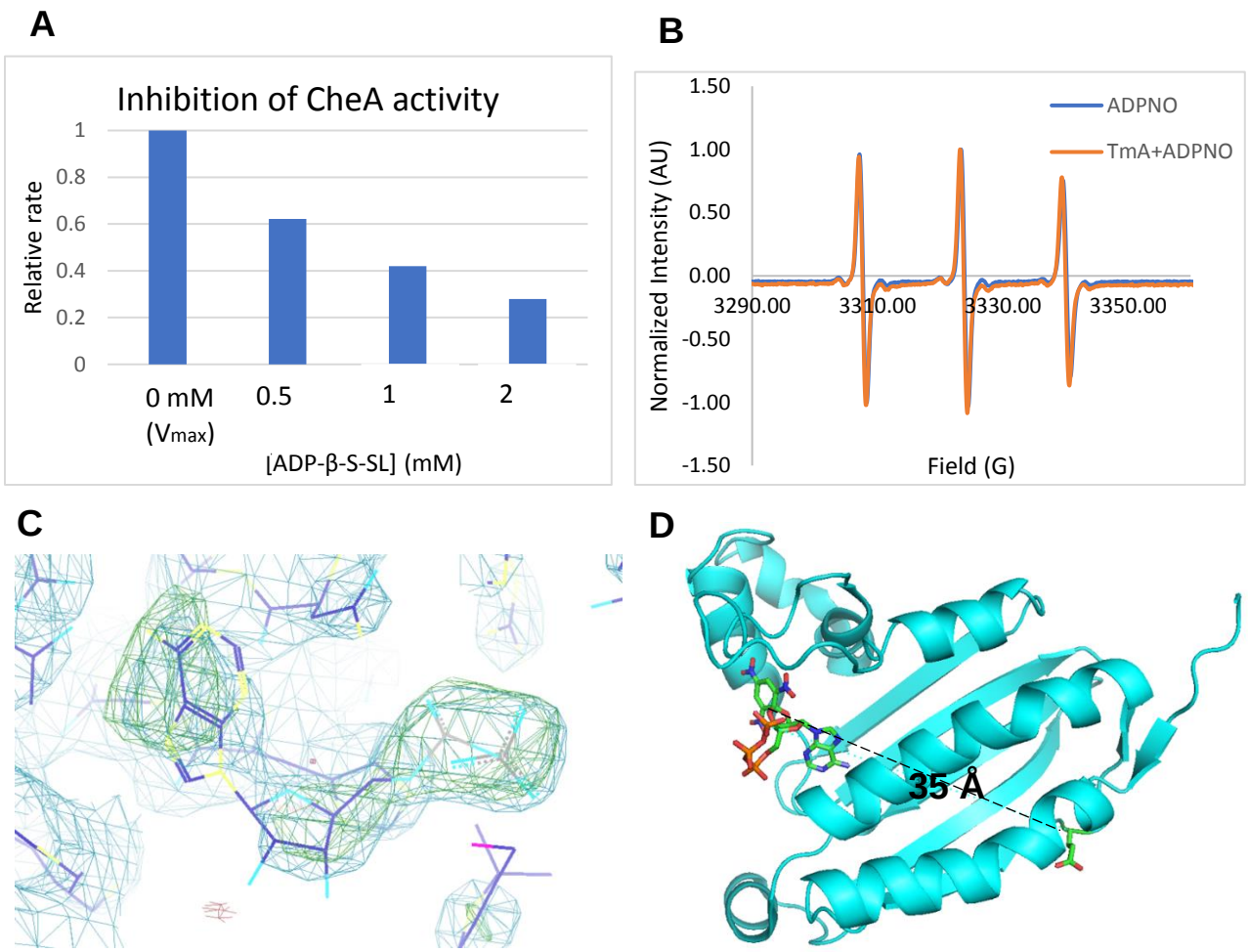


Figure 2: (A) The relative velocity for CheA autophosphorylation with different concentrations of ADP-β-S-SL. (B) Continuous wave ESR of free ADP-β-S-SL and ADP-β-S-SL reconstituted into CheA demonstrate that the nitroxide is highly mobile when bound in the CheA nucleotide pocket. (C) The crystal structure of the CheA nucleotide binding domain (P4) reconstituted with ADP-β-S-SL reveals density for the synthetic nucleotide. However, as the nitroxide label is highly mobile, there is only obvious density corresponding to the ADP moiety.

successfully utilized in ESR experiments. Using a similar approach, we herein present a spin labeling method for ATP-binding proteins that adds a label directly to the substrate nucleotide. Therefore, this method allows for direct labeling of the nucleotide-binding region without

perturbing protein function. The feasibility of this approach is demonstrated with the ATP-binding histidine kinase CheA, a flexible protein that contains several domains connected by long flexible linkers. Although this method is presented specifically for ATP/ADP-binding proteins, this approach can also readily be applied to GTP/GDP-binding proteins.

4.2 Results

4.2.1 MTSSL labeling and isolation of ADP-β-S-SL

The ESR spin label MTSL was coupled to the ADP analog ADP-β-S by reaction of ADP-β-S with MTSL in aqueous solution for 10 hr (Fig1 A). The product, ADP-β-S-SL, was readily detected by thin-layer chromatography and then separated from the precursors and methanesulfonyl byproduct by liquid-chromatography mass-spectrometry (LC-MS)(Fig. 1B). Solid ADP-β-S-SL was then isolated by high-performance liquid chromatography (HPLC) followed by rotary evaporation. Subsequent analysis of the isolated product by LC-MS demonstrates that ADP-β-S-SL is present in high purity and stable in solution after freezing, storage at -80°C, and 24 hours incubation at room temperature. LC-MS experiments identify adenosine monophosphate (AMP) as the degradation product of ADP-β-S-SL (Fig. 1C).

4.2.2 Binding of ADP-β-S-SL to CheA

To confirm that isolated ADP-β-S-SL binds in the ATP-binding pocket of *T. maritima* CheA, radioisotope assays using ADP-β-S-SL as an inhibitor for CheA autophosphorylation were conducted. ATP binds to the Tm CheA dimer with a KD of ~1 µM. ADP-β-S-SL acts as a competitive inhibitor for ATP binding with an apparent Ki ~ 29 µM (Fig. 2A).

The ATPase domain of CheA (P4) was co-crystallized with ADP-β-S-SL and the structure determined to 2.9 Å resolution by x-ray crystallography. The structure of CheA[ADP-β-S-SL] reveals a P4 homo-dimer similar to that observed in the structure of CheA P4 with the fluorescent ATP analog TNP-ATP[5]. ADP-β-S-SL binds analogously to ADP with a key hydrogen

bond formed between Asp449 and the adenine ring at the base of the P4 β-sheet. Typical of diphosphate nucleotide binding, the P4 ATP-lid region is open and not engaged with divalent metal ion. Although the phosphate moieties of ADP are well defined, there is very little discernible electron density for the nitroxide moiety itself, which resides at the surface of the

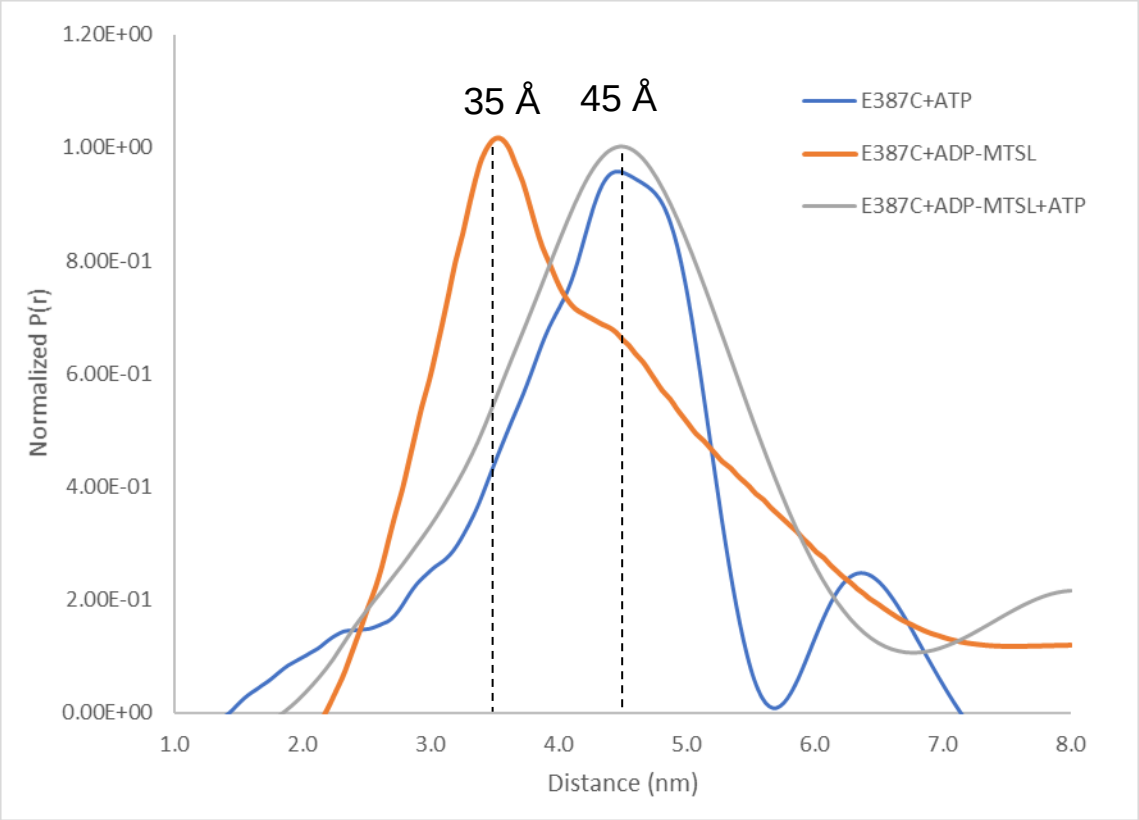


Figure 3: ESR DEER experiments demonstrate that CheA labeled on the P4 domain at position E387C produces a distance distribution with a broad peak centered at 45 Å. When ADP-β-S-SL is reconstituted into the protein, a peak centered at 35 Å is evident. Furthermore, when excess ATP is added to the sample to out-compete ADP-β-S-SL from the nucleotide pocket, the 35 Å peak is diminished.

protein and appears disordered.

4.2.3 ESR measurements of ADP-β-S-SL bound to CheA

Cw-ESR at 25 °C was carried with full-length CheA after incubation with stoichiometric amounts of ADP-β-S-SL. The resulting spectra are indicative of freely mobile spin-label, which is consistent with the crystallographic results (Fig 2B).

Pulsed ESR spectroscopy demonstrates that the isolated ADP-β-S-SL has a strong primary echo signal that decays in a manner similar to MTSSL. To demonstrate that ADP-β-S-SL can be used as a label for DEER measurements, the purified compound was incubated with full-length CheA containing an MTSL-labeled cysteine at residue 387 (E387C-SL). CheA is a dimeric kinase with each subunit (MW of 79K), containing five domains connected by flexible linkers. Without ADP-β-S-MTSL present, the E387C-SL CheA produces a distance distribution (P(r)) with a peak centered at 45 Å, which represents the distance between the active sites of the P4 domains in the dimer[6]. However, when ADP-β-S-SL is added to E387C-SL a shorter distance of about 35 Å becomes apparent (Fig. 3). This shorter separation agrees with the expected distance between the E387C-SL and ADP-β-SL bound to P4 in the crystal structure (Fig. 2D). Upon addition of excess (5mM) ATP the short distance is lost from the distribution.

4.3 Discussion

In principle, any ATP binding protein would be a suitable subject for spin-labeling by this strategy. That said, in some cases the nitroxide modification may interfere with nucleotide binding. Nonetheless, the technique is straightforward and can provide immediate distance information on complex molecules without further genetic engineering. Many ATP-binding proteins are oligomers and hence this method can report on their assembly state as well as active site dispositions. Combined with Cys-directed spin-labeling, both internal distance constraints and intermolecular constraint to interacting partner proteins are readily accessible.

4.4 Methods and Materials

4.4.1 Synthesis of ADP-β-S-SL

ADP-β-S and MTSSL were mixed in a 1:2 molar ratio in a solution containing 50 mM potassium phosphate buffer pH 7.5 and 20% acetonitrile. The reaction was allowed to proceed at room temperature for 3 hours after covering the tube with aluminum foil. The reaction mixture was then injected on an HPLC column (Agilent Eclipse, XDB-C18, 5 µM 9.4 X 250 mm) pre-equilibrated with water and eluted with an increasing gradient of acetonitrile (0-20%). The fractions containing ADP-β-S-MTSL were dried by rotary evaporation. The isolated ADP-β-S-SL was then resuspended in 50 mM Potassium phosphate buffer pH 7.5 and 20% acetonitrile, separated into aliquots, flash frozen, and stored at -80°C.

4.4.2 Purification of *T. maritima* CheA

CheA from *T. maritima* was cloned from genomic DNA into a pet28a vector that contains an N-terminal poly-His tag and a kanamycin resistance marker. pet28a was transformed into *E. coli* BL21 competent cells. The cells were grown in 100 ml of LB at 37 °C overnight and then transferred into 8L of LB. The cells were grown at 37 °C to an O.D. of 0.8 and then induced by the addition 100 µM IPTG at 25 °C. The cells were collected by centrifugation after 16 hours of protein growth and lysed via sonication. The lysate was subjected to centrifugation to pellet insoluble cellular components. The soluble portion of the lysate was run over a column containing Ni-NTA. The resin was washed with buffer to remove all proteins except for the His-tagged CheA. Protein was subjected to SEC and concentrated via high-speed centrifugation. CheA P4 (residues 350-540), was also expressed from pet28a and purified using a similar procedure as above. For both proteins, the poly-His tag was removed via protein cleavage with thrombin.

4.4.3 MTSL labeling of CheA

To add the MTSL label to CheA E387C, 5 mg of MTSSL was dissolved in 100ul of acetonitrile and added to 5ml of buffer. The 5 ml of buffer was added to Ni-NTA resin containing bound poly-His CheA and incubated for 16 hours at 4C. The unreacted MTSSL was washed from the resin with buffer and the protein was eluted from the Ni-NTA with imidazole. The eluted protein was further purified using an s200 GE-Sephadex column.

4.4.4 Crystallization of CheA P4 with ADP-β-S-SL

The P4 (ATP-binding) domain of CheA was co-crystallized with ADP-β-S-MTSL via hanging drop vapor diffusion by mixing 2 µl of 1 mM P4 solution (50 mM Tris pH 7.5, 150 mM NaCl) containing 3 mM ADP-β-S-SL with 2 µl of crystallization buffer (50 mM Citrate pH 5.5, 2 M Ammonium sulfate). The rectangular crystals formed within 24 hours and diffracted to about 2.9

Table 4.1 Data collection and refinement statistics for P4(ADP-β-S-SL)

Wavelength (Å)	0.97220
Synchrotron	APS
Beam-line	24-ID-E
Space group	$P2_1$
a, b, c	83.703, 75.145, 86.580
α, β, γ	90, 90, 90
Resolution (Å)	14.8 – 2.9
R_{merge}	0.032(0.875)
$R_{p.i.m.}$	0.040(0.755)
R_{meas}	0.071(1.341)
I/σ(I)	5.7 (1.0)
Completeness (%)	95.8(95.47)
Multiplicity	3.0(2.0)
Mosaicity	0.475
Total reflections	25447
Refinement	
Resolution (Å)	2.9
Reflections used for R_{free}	22833
R_{free}	0.3076
R_{work}	0.2573
Clash Score	40.24
R.M.S Deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.90
Ramachandran outliers (%)	0.14
Rotamer outliers (%)	21.95

Å. The protein crystallized as a homo-dimer in a P 2_1 space group. The density for ADP-β-S-SL is well occupied in only one of the protein subunits.

4.4.5 Electron-spin resonance spectroscopy

For Pulsed Dipolar Spectroscopy (PDS) measurements, four pulse double electron electron resonance (DEER) experiments were conducted at 60 K on a 17.3 GHz FT EPR spectrometer, which is modified to perform PDS experiments. The baseline used for data processing was approximated by a linear polynomial. Distance distributions of spin separations within the sample were calculated by the Tikhonov method[7].

The cw-ESR spectra were obtained on a Bruker E500 ESR spectrometer operating at X-band (~9.4 GHz). The ESR measurements were carried out at room temperature with a modulation amplitude of 0.5 G and modulation frequency of 100 kHz. All spectra were digitized to 1024 points with a sweep width of ~100 G and an average of 4-16 scans.

4.5 References

1. Klare, J. P. Site-directed spin labeling EPR spectroscopy in protein research. **394,** 1281–1300 (2013).

2. Schmidt, M. J., Borbas, J., Drescher, M. & Summerer, D. A Genetically Encoded Spin Label for Electron Paramagnetic Resonance Distance Measurements. **3,** (2014).

3. Bertrand, C. *et al.* The Use of Mn(II) Bound to His-tags as Genetically Encodable Spin-Label for Nanometric Distance Determination in Proteins. (2016). doi:10.1021/acs.jpclett.6b00362

4. Grant, G. P. G. & Qin, P. Z. A facile method for attaching nitroxide spin labels at the 5 0 terminus of nucleic acids y. **35,** 1–8 (2007).

5. Bilwes, A. M., Quezada, C. M., Croal, L. R., Crane, B. R. & Simon, M. I. Nucleotide binding by the histidine kinase CheA. *Nat. Struct. Biol.* **8,** 353–360 (2001).

6. Bhatnagar, J. *et al.* Structure of the ternary complex formed by a chemotaxis receptor signaling domain, the CheA histidine kinase, and the coupling protein CheW as determined by pulsed dipolar ESR spectroscopy. *Biochemistry* **49,** 3824–41 (2010).

7. Chiang, Y., Borbat, P. P. & Freed, J. H. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. **172,** 279–295 (2005).

Appendix A

DSSO crosslinks identified in free *Thermotoga maritima* CheA and a CheA:CheW:Tar foldon complex

1.) Free Tm CheA(ADP)

Domains	Residue A	Residue B	PDB	PDB Distance	Sequence A	Sequence B
P2 to P2	205	195	1U0S	15.5 Å	H[K]LEEL	[K]SAR
P3 to P5	328	651	1B3Q	45.9 Å	[K]KY	SEV[K]EF
P4 to P4	469	499	2CH4	17.4 Å	KAIE[K]GLIDESK	STKE[K]VSEVSGR
P2 to P2	205	210	1U0S	9.8 Å	H[K]LEEL	[K]CEVVR
P4 to P4	461	499	2CH4	21.9 Å	HEGNNVVIEVEDDGRGIDKE[K]IIRK	STKE[K]VSEVSGR
P2 to P2	210	177	1U0S	12 Å	[K]CEVVR	N[K]GFK
P4 to P4	499	465	2CH4	18.2 Å	E[K]VSEVSGR	[K]AIEK
P5 to P3	607	333	1B3Q	39.4 Å	QIEH[K]EELEEMEAVIVR	NI[K]ELDESLSHL
P5 to P5	607	553	1B3Q	12.2 Å	EVLQIEH[K]EELEEMEAVIVR	V[K]VNNLVY
P3 to P3	333	329	1B3Q	11.5 Å	NI[K]ELDESLSHL	ILETLK[K]Y
P4 to P4	421	499	2CH4	30 Å	NAIDHGIEP[K]EER	STKE[K]VSEVSGR
P4 to P4	461	427	1B3Q	22 Å	HEGNNVVIEVEDDGRGIDKE[K]IIR	IA[K]GKPPIGTL
P5 to P5	607	574	1B3Q	15.2 Å	EVLQIEH[K]EELEEMEAVIVR	SIS[K]EDIQRVQDRDVIVIRGEVIPVY
P4 to P4	459	465	1B3Q	10 Å	SARHEGNNVVIEVEDDGRGID[K]EKIIR	[K]AIEK
P1 to P1	90	75	1TQG	8.7 Å	D[K]IF	CHTLENILD[K]AR
P1 to P1	64	75	1TQG	17 Å	SSMA[K]L	CHTLENILD[K]AR
P4 to P4	461	465	1B3Q	16.6 Å	HEGNNVVIEVEDDGRGIDKE[K]IIR	[K]AIEK
P4 to P4	421	377	1B3Q	14.4 Å	RNAIDHGIEP[K]EER	DLA[K]K
P4 to P4	499	378	2CH4	25.2 Å	STKE[K]VSEVSGR	[K]MNKEVNF
P2 to P2	205	177	1U0S	20.6 Å	H[K]LEEL	N[K]GF
P3 to P3	333	328	1B3Q	11 Å	ILETL[K]ELDESLSHL	ETL[K]KY
P4 to P4	459	499	2CH4	23.6 Å	HEGNNVVIEVEDDGRGID[K]EKIIRK	STKE[K]VSEVSGR
P4 to P5	476	624	1B3Q	24 Å	KAIEKGLIDES[K]AATL	VGNR[K]Y
P1 to P1	125	75	2LD6	8.7 Å	SDTI[K]SF	CHTLENILD[K]AR
P4 to P4	459	427	1B3Q	18.9 Å	SARHEGNNVVIEVEDDGRGID[K]EKIIR	IA[K]GKPPIGTL
P3 to P5	333	651	1B3Q	44.8 Å	NI[K]ELDESLSHL	SEV[K]EF
P2 to P2	185	195	1U0S	14.6 Å	I[K]VILKEGTQL	[K]SAR
P5 to P5	641	645	1B3Q	19 Å	GIVVDDLLGQDDIVI[K]SL	G[K]VF
P4 to P5	476	651	1B3Q	24.4 Å	GLIDES[K]AATL	SEV[K]EF
P3 to P5	333	553	1B3Q	27.9 Å	NI[K]ELDESLSHL	V[K]VNNLVY
P5 to P5	651	645	1B3Q	14.3 Å	SEV[K]EF	SLG[K]VF

2.) Complex of TmCheA:TmCheW:Tar foldon

Domains	Position A	Position B	PDB file	PDB distance	Sequence A	Sequence B
R to W	159	80	-	-	SAQAA[K]EIK	ILGISFDEQ[K]MK
P1 to P1	125	82	1TQG	17.6 Å	SDTI[K]SF	NSEI[K]ITSDLLDK
W to W	39	82	3UR1	4.8 Å	AFDVDNIEMVIE[K]SDITPVPK	M[K]SIIVAR
P1 to P4	90	378	-	-	ITSDLLD[K]IF	[K]MNK
P2 to P4	499	189	-	-	E[K]VSEVSGR	[K]EGTQLK
P1 to P4	90	499	-	-	ITSDLLD[K]IF	E[K]VSEVSGR
W to W	120	70	3UR1	13.9 Å	ITENQLDLTNVSD[K]F	A[K]IL
W to W	39	70	3UR1	15.5 Å	AFDVDNIEMVIE[K]SDITPVPK	IIPVVNLA[K]IL
W to W	151	123	3UR1	18 Å	LDIDKIIEEITV[K]EGV	G[K]K
P2 to P2	244	205	1U0S	13 Å	VISPVDLE[K]LSEAL	LVFH[K]LEEL
R to R	103	104	-	-	ALIEDSVSGSAGTSGDSS[K]K	[K]IADIISVIDGIAF
P2 to P2	244	180	1U0S	13.5 Å	VISPVDLE[K]LSEAL	GF[K]TFY
W to P5	151	651	3UR1	14 Å	LDIDKIIEEITV[K]EGV	VFSEV[K]EF
P1 to P1	143	90	1TQG	15 Å	LDID[K]IIEEITVK	T[K]DVEVGF
P4 to P4	499	378	1B3Q	28.1 Å	E[K]VSEVSGR	[K]MNK
W to W	124	120	3UR1	10 Å	[K]SK	ITENQLDLTNVSD[K]FGK
P2 to W	189	130	-	-	[K]EGTQLK	V[K]TDGR
P2 to W	210	130	-	-	[K]CEVVR	V[K]TDGR
W to P1	130	64	-	-	V[K]TDGR	SSMA[K]L
R to P2	159	189	-	-	SAQAA[K]EIK	[K]EGTQLK
P3 to W	293	90	-	-	[K]VISQTVR	T[K]DVEVGF
P4 to P4	532	427	3UR1	10.8 Å	NVVESLNGSISIESEKD[K]GTK	IA[K]GKPPIGTL
P4 to P4	459	427	2CH4	18.4 Å	HEGNNVVIEVEDDGRGID[K]EKIIR	IA[K]GKPPIGTL
W to W	151	90	3UR1	22 Å	LDIDKIIEEITV[K]EGV	T[K]DVEVGF
W to P4	151	476	-	-	LDIDKIIEEITV[K]EGV	GLIDES[K]AATL
R to R	159	25	-	-	SAQAA[K]EIK	GSHMDSS[K]K
P2 to P2	244	210	1U0S	13.3 Å	VISPVDLE[K]LSEAL	LEEL[K]CEVVR
P3 to P3	333	328	1B3Q	20 Å	YNI[K]ELDESLSHL	ILETL[K]K
P1 to P4	32	421	-	-	ELE[K]NPEDMELINEAFR	NAIDHGIEP[K]EER
R to W	159	130	-	-	SAQAA[K]EIK	V[K]TDGR
P4 to P4	499	465	1B3Q	13 Å	E[K]VSEVSGR	[K]AIEK
P1 to P2	130	177	-	-	V[K]TDGR	N[K]GF
P2 to P2	210	177	1U0S	14 Å	[K]CEVVR	N[K]GFK
R to R	103	159	-	-	ALIEDSVSGSAGTSGDSS[K]K	SAQAA[K]EIK
P4 to P4	499	461	1B3Q	17.6 Å	E[K]VSEVSGR	HEGNNVVIEVEDDGRGIDKE[K]IIR
P4 to P3	421	377	3UR1	14.2 Å	NAIDHGIEP[K]EER	DLA[K]K
W to W	120	130	3UR1	18.8 Å	ITENQLDLTNVSD[K]FGK	GLV[K]TDGR
P1 to P1	125	75	1TQG	5 Å	SDTI[K]SF	LCHTLENILD[K]AR
P4 to P4	461	427	2CH4	16 Å	HEGNNVVIEVEDDGRGIDKE[K]IIR	IA[K]GKPPIGTL
R to P3	159	328	-	-	SAQAA[K]EIK	ILETL[K]K
R to W	159	82	-	-	SAQAA[K]EIK	M[K]SIIVAR
R to R	159	193	-	-	SAQAA[K]EIK	[K]DGEW
P4 to W	476	123	-	-	GLIDES[K]AATL	G[K]K
W to P5	47	651	3UR1	8.7 Å	SDITPVP[K]SR	SEV[K]EF
W to W	47	70	3UR1	19.6 Å	SDITPVP[K]SR	A[K]IL
P3 to R	333	159	-	-	YNI[K]ELDESLSHL	SAQAA[K]EIK

Domains	Position A	Position B	PDB file	PDB distance	Sequence A	Sequence B
P5 to P5	607	574	3UR1	14.6 Å	QIEH[K]EELEEMEAVIVR	SIS[K]EDIQR
P4 to P4	459	499	2CH4	30 Å	HEGNNVVIEVEDDGRGID[K]EKIIR	E[K]VSEVSGR
P4 to P4	469	499	2CH4	10 Å	AIE[K]GLIDESK	E[K]VSEVSGR
P1 to P1	75	90	1TQG	8 Å	ENILD[K]AR	ITSDLLD[K]IF
P2 to P2	189	195	1U0S	12 Å	[K]EGTQL	[K]SAR
P5 to P4	530	429	3UR1	11 Å	NVVESLNGSISIESE[K]DKGTK	IAKG[K]PPIGTLIL
W to P5	151	645	3UR1	5 Å	LDIDKIIEEITV[K]EGV	G[K]VF
P1 to P2	90	244	-	-	D[K]IF	VISPVDLE[K]LSEAL
P2 to P1	244	64	-	-	VISPVDLE[K]LSEAL	SSMA[K]L
P4 to W	514	90	3UR1	18.3 Å	GVGMDVV[K]NVVESL	T[K]DVEVGF
P2 to W	244	130	-	-	VISPVDLE[K]LSEAL	GLV[K]TDGR
P4 to W	476	47	-	-	GLIDES[K]AATLSDQEIL	SDITPVP[K]SR
P1 to P4	90	421	-	-	D[K]IF	NAIDHGIEP[K]EER
P4 to P4	530	427	3UR1	12 Å	NVVESLNGSISIESE[K]DKGTK	IA[K]GKPPIGTL
P4 to P5	476	651	1B3Q	17.4 Å	GLIDES[K]AATL	VFSEV[K]EF
P2 to P4	180	499	-	-	[K]TF	E[K]VSEVSGR
P4 to P5	514	651	3UR1	24.2 Å	GVGMDVV[K]NVVESLNGSISIESEKDK	VFSEV[K]EF
P2 to P4	244	499	-	-	VISPVDLE[K]LSEAL	E[K]VSEVSGR
P5 to P5	641	645	3UR1	11.6 Å	GIVVDDLLGQDDIVI[K]SL	G[K]VF
P2 to P4	180	421	-	-	[K]TF	NAIDHGIEP[K]EER
P2 to P4	421	177	-	-	NAIDHGIEP[K]EER	N[K]GF
P1 to P2	125	189	-	-	SDTI[K]SF	VIL[K]EGTQLK
W to W	151	47	3UR1	17 Å	LDIDKIIEEITV[K]EGV	SDITPVP[K]SR

Appendix B

The P4 domain of CheA possesses ATPase activity

Previous research in the Crane lab (Greenswag et al.) indicates that phosphorylated full-length *Thermotoga maritima* (Tm) CheA rapidly dephosphorylates at 55°C. To experimentally determine the concentration of inorganic phosphate (Pi) over time, a colorimetric assay that monitors Pi was incubated with CheA and ATP at 55°C. The assay utilizes a reagent mixture called BioMol Green that contains the necessary components to form malachite green, which absorbs at 640 nm (Fig. 1A). As expected, the concentration of Pi increases over time (Fig. 1B).

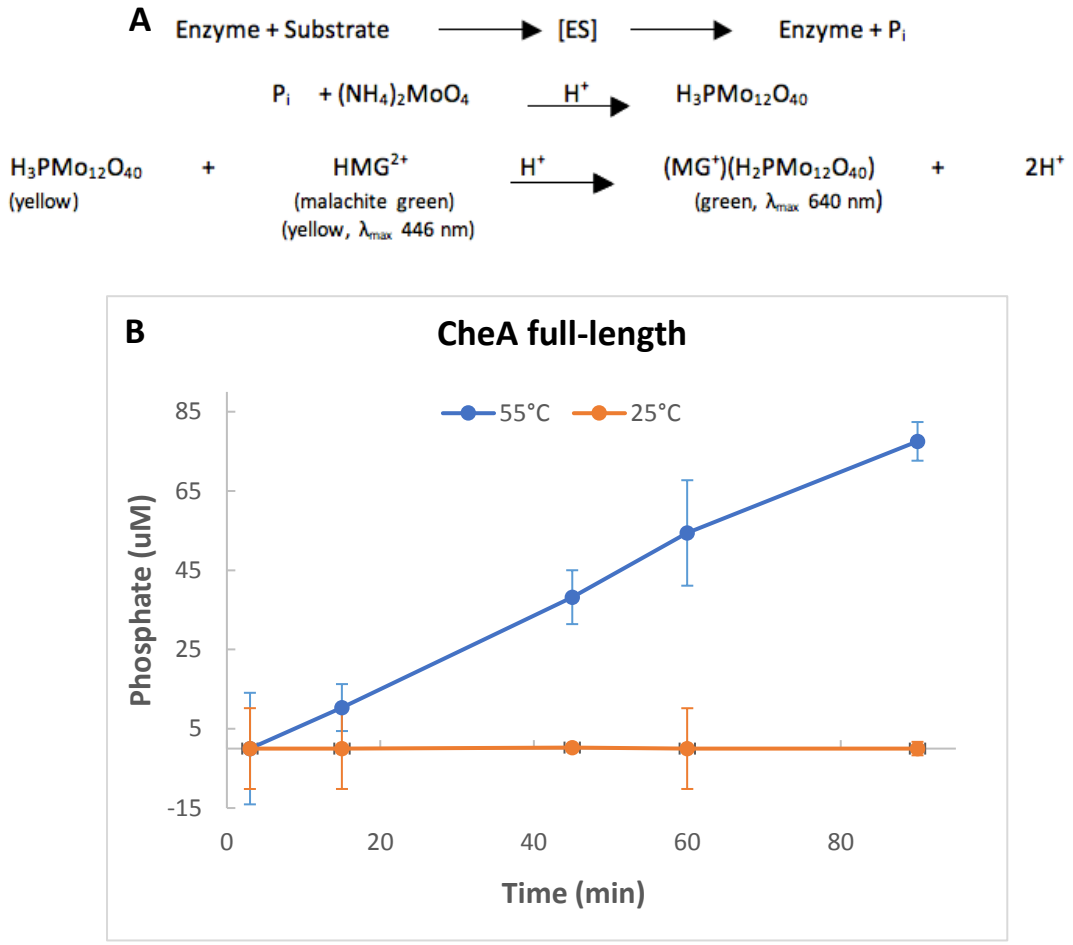


Figure 1: (A) The reaction scheme for the addition of the BioMol Green reagent mixture to a solution containing inorganic phosphate. (B) The time course assay for the release of inorganic phosphate from full-length CheA at 25°C and 55°C.

To confirm that the presence of Pi is due to dephosphorylation of the P1 domains exclusively, identical experiments were carried out with a H45K CheA mutant and a variant of Tm CheA that does not possess the P1 or P2 domains (Δ289); both variants are incapable of autophosphorylation. Unexpectedly, both variants also produced a significant amount of phosphate (Fig. 2). Furthermore, while the full-length H45K mutant produces similar amounts of Pi as WT CheA, the Δ289 variant produces significantly more. These data suggest that Tm CheA acts as an ATPase and that the P1 domains can inhibit this activity.

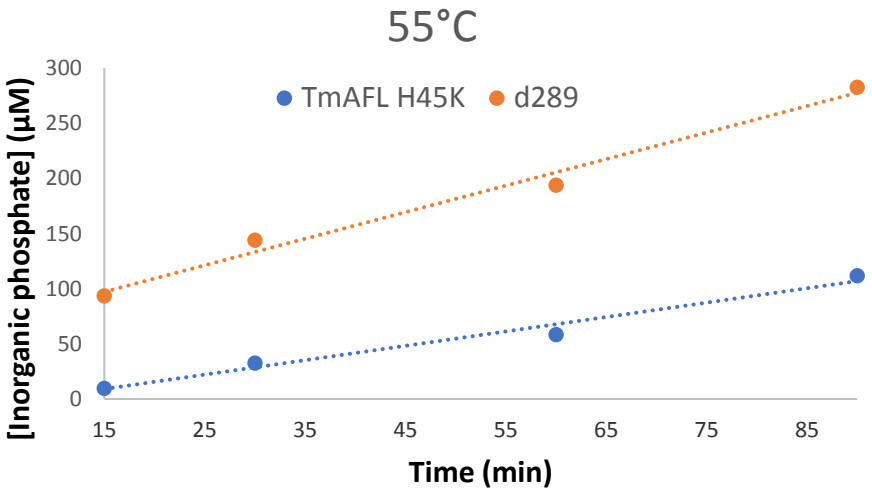


Figure 2: Inorganic phosphate assays with Tm CheA H45K and Δ289 demonstrate that Tm CheA is an ATPase. Furthermore, Δ289 produces more phosphate than the full-length kinase.

BioMol green assays were conducted with just the isolated P4 domain to determine if this domain on its own is responsible for the production of Pi. As a negative control, the experiment was also conducted with the isolated P5 domain. The results demonstrate that the presence of the inorganic phosphate is due to the P4 domain interacting with ATP. The P5 domain does not possess any ATPase activity (Fig. 3).

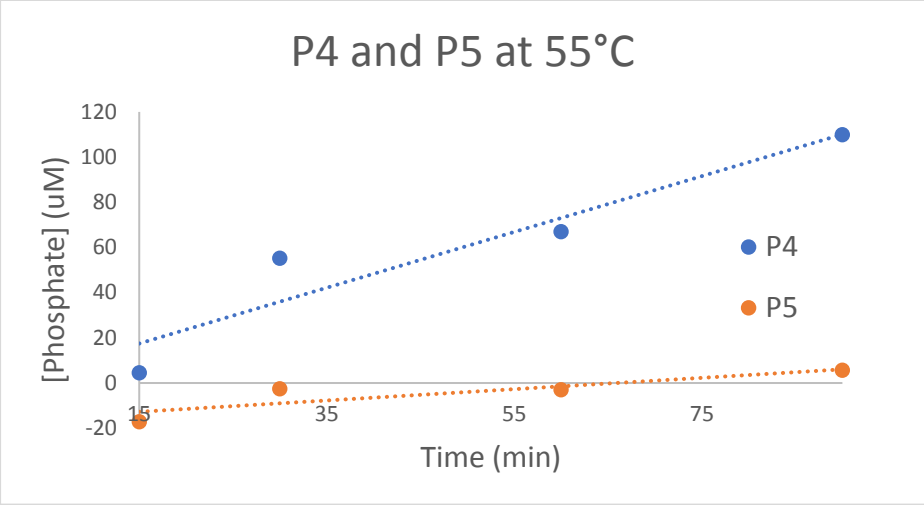


Figure 3: Inorganic phosphate assays with CheA isolated P4 and P5 domains demonstrate that the P4 domain produces inorganic phosphate when incubated with ATP.

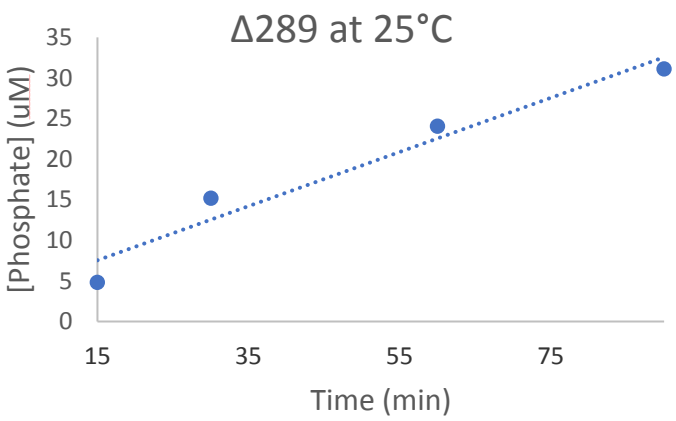


Figure 4: Inorganic phosphate assays with Δ289 at 25°C demonstrate that the P1/P2 domains inhibit ATPase activity at this temperature.

To deduce whether the P1/P2 domains of Tm CheA do inhibit this activity, BioMol green assays were carried out with Δ289 at 25°C; for full-length TmA Pi is not produced at 25°C. Interestingly, at lower temperatures the Δ289 variant does possess ATPase activity (Fig. 4).

Identical experiments were then conducted with *Escherichia coli* (Ec) to determine if the ATPase activity is conserved across bacterial species. Experiments were conducted with the full-length kinase and a variant that lacks the P1 and P2 domains (Δ258) at different temperatures. As seen with Tm CheA, Ec Δ258 possess ATPase activity as the temperature increases. Unfortunately the Δ258 variant is not stable at 37°C, which is the physiological

temperature of *E. coli* cells, and precipitates during the course of the experiment. The results of these experiments demonstrate that Ec CheA also possess ATPase activity at near-physiological temperatures but only in the absence of the P1/P2 domains (Fig. 5).

Further experiments were done to see if the ATPase activity of either Ec or Tm CheA was mediated by receptor interactions but the results were inconclusive due to the instability of purified receptors. Therefore, further experimentation must be done to identify the physiological importance of this activity.

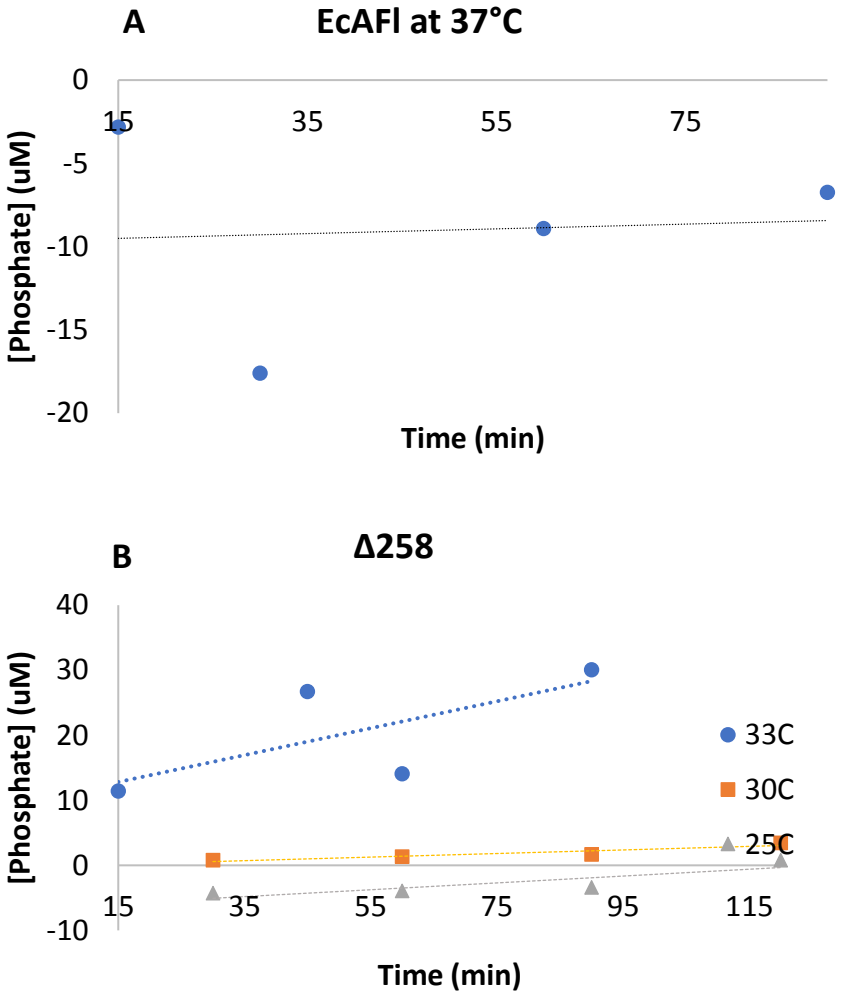


Figure 5: (A) Inorganic phosphate assays with Ec CheA full length at 37°C. (B) Phosphate assays with Ec Δ258 variant at different temperatures. Variant precipitates at 37°C.